

For Reference

NOT TO BE TAKEN FROM THIS ROOM

Ex libris
UNIVERSITATIS
ALBERTAENSIS



THE UNIVERSITY OF ALBERTA

RELEASE FORM

NAME OF AUTHOR Sherif Rushdy
TITLE OF THESIS Maximum Likelihood Estimation of a
 Nonlinear System Dynamic Market Growth
 Model

DEGREE FOR WHICH THESIS WAS PRESENTED Master of Science

YEAR THIS DEGREE GRANTED October 1980

Permission is hereby granted to THE UNIVERSITY OF ALBERTA LIBRARY to reproduce single copies of this thesis and to lend or sell such copies for private, scholarly or scientific research purposes only.

The author reserves other publication rights, and neither the thesis nor extensive extracts from it may be printed or otherwise reproduced without the author's written permission.

THE UNIVERSITY OF ALBERTA

Maximum Likelihood Estimation of a Nonlinear System Dynamic
Market Growth Model

by



Sherif Rushdy

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES AND RESEARCH
IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE
OF Master of Science

Electrical Engineering

EDMONTON, ALBERTA

SPRING 1981

61-99

THE UNIVERSITY OF ALBERTA
FACULTY OF GRADUATE STUDIES AND RESEARCH

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research, for acceptance, a thesis entitled Maximum Likelihood Estimation of a Nonlinear System Dynamic Market Growth Model submitted by Sherif Rushdy in partial fulfilment of the requirements for the degree of Master of Science.

Dedication

To my parents for everything

Abstract

Recent efforts to determine the proper role of formal statistical estimation in modeling "System Dynamics" models show that the parameter estimates derived from Ordinary and Generalized Least Squares (OLS and GLS) are highly sensitive to errors in data measurement, and likely to prove misleading if used as a basis for selection of parameter values or structural analysis.

Using the framework developed in this previous work, - where a nonlinear feedback model generates synthetic data, which is then used to estimate model parameters and thus provide a basis for the evaluation of an estimation technique, - this thesis reviews previous results with OLS and investigates alternative estimation techniques.

A review of both econometric and engineering techniques, together with some preliminary experimental results revealed that no econometric estimation technique proved capable of meeting the requirements of the estimation of parameters in a nonlinear dynamic feedback model in the presence of measurement noise.

The only promising method, the Filtering Form of the Maximum Likelihood algorithm, was found in the engineering literature where it is being used in a growing number of applications.

A general FORTRAN program was developed to implement this algorithm and was tried out on three small-scale linear and nonlinear models. The method was found to be capable of

drastically improving upon the Least Squares estimates, if sufficient knowledge about the noise statistics (which present identifiability problems if they are all to be estimated) was available.

However, experimentation on Forrester's "Market-Growth" model, while still modest in size compared to many System Dynamics models (nine equations and fifteen parameters to estimate), revealed the many limitations (in particular in convergence and cost) of this algorithm, that preclude its use in socio-economic applications.

In the light of the above results, alternative methods of model validation, and in particular a more formal use of sensitivity testing, are suggested for further research.

Acknowledgements

I would like here to express my gratitude to Dr. Rink who introduced me to the whole field of System Modeling and Identification, provided me with the opportunity to materialize some of my interests and the flexible supervision and encouragement that permeated this endeavour; to Dr. Buse who familiarized me with the intricacies of Econometric theory and whose notation and explanations reappear in many pages of this thesis.

Financial support from the University of Alberta and the National Research council is also gratefully acknowledged.

Table of Contents

Chapter	Page
I. INTRODUCTION	1
A. THE NEED OF OUR TIMES	1
B. TWO APPROACHES TO MODELING	2
A.1 Econometric models	2
A.2 System dynamics models	5
C. SENGE'S ATTEMPT	15
D. THIS THESIS	16
II. SYSTEM IDENTIFICATION: AN OVERVIEW	20
III. SOME ECONOMETRIC ESTIMATION TECHNIQUES	26
A. NOTATIONS AND ASSUMPTIONS	26
B. SOME ESTIMATION TECHNIQUES	28
B.1 Desirable properties for an estimator ..	28
B.2 Single equation methods	29
B.3 Systems methods	33
C. DEPARTURES AND CURRENT RESEARCH AREAS	34
C.1 Multicollinearity	35
C.2 Autocorrelation	35
C.3 Time-varying parameters	36
C.4 Errors in the variables	36
C.5 Small sample properties	40
C.6 Estimation under Nonlinear Specification	42
D. THE LISREL MODEL	43
IV. THE ENGINEERING LITERATURE	47
A. NOTATION	47

B.	SOME ESTIMATION TECHNIQUES	50
B.1	Bayes Estimator	50
B.2	Maximum Likelihood Estimator (MLE)	51
B.3	Markov and Least Squares Estimates	53
C.	CURRENT RESEARCH	55
C.1	Model Validation - Sensitivity Testing	56
C.2	The use of Information Theory	57
C.3	Case Studies	59
V.	THE IDENTIFIABILITY PROBLEM	60
VI.	EXPERIMENTAL WORK	66
A.	INTRODUCTION	66
B.	THE EXPERIMENTAL FRAMEWORK	68
B.1	The experimental procedure	68
B.2	Variations in experimental conditions	69
VII.	REPRODUCING SENGE'S EXPERIMENTS	73
A.	THE "MARKET GROWTH" MODEL	73
B.	EXPERIMENTAL RESULTS	79
B.1	Ideal conditions	79
B.2	Imperfections in data Measurement	90
VIII.	ALTERNATIVE ECONOMETRIC METHODS	99
A.	INTRODUCTION	99
B.	LISREL RUNS	101
B.1	The Identifiability Problem	101
B.2	Experimental results	101
IX.	THE FILTERING FORM OF THE LIKELIHOOD FUNCTION ...	103
A.	PHILOSOPHY OF THE MAXIMUM LIKELIHOOD METHOD ..	103
B.	MAXIMUM LIKELIHOOD AND IDENTIFICATION	104

C.	AN INTUITIVE APPROACH TO THE FILTERING FORM OF THE ML ALGORITHM	110
C.1	OLS	110
C.2	LIKALM	112
D.	DERIVATION OF THE LIKALM ALGORITHM FOR A ONE-DIMENSIONAL MODEL	116
D.1	Statement of the problem	116
D.2	Solution of the problem	117
D.3	Unbiasedness of the LIKALM estimate ...	123
E.	THE KALMAN FILTER	125
E.1	Statement of the problem	125
E.2	Solution of the problem	127
E.3	The Extended Kalman filter	129
E.4	A word on innovations	130
F.	DERIVATION OF THE FILTERING FORM OF THE ML ESTIMATOR	131
F.1	Expressions of the gradient and the information matrix	135
F.2	Recursive equations	137
X.	ESTIMATION WITH THE FILTERING FORM OF THE MAXIMUM LIKELIHOOD	140
A.	THE LIKALM COMPUTER PROGRAM	140
A.1	Main features of the LIKALM program ...	140
A.2	The FORTRAN equations	143
B.	EXPERIMENTAL RESULTS	146
B.1	One Dimensional Model	146
B.2	Two Dimensional Oscillator	169
B.3	Three-Dimensional Nonlinear Model	188
B.4	The Market-Growth model	195

XI. CONCLUSION	214
A. Introduction	214
B. Limitations of the LIKALM algorithm	216
C. Suggestions for Further Research	220
REFERENCES	223

List of Tables

Table	Page
VII- 1. Market-Growth model symbols.....	78
VII- 2. Experimental conditions for Tables VII-3 to VII-6 (Ideal situation: 1% equation noise).....	82
VII- 3. Market model - Ideal conditions: OLS estimates of parameters in equation 1.....	85
VII- 4. Market model - Ideal conditions: OLS estimates of parameters in equation 2.....	86
VII- 5. Market model - Ideal conditions: OLS estimates of parameters in equation 3.....	87
VII- 6. Market model - Ideal conditions: OLS estimates of parameters in equation 4.....	88
VII- 7. Experimental conditions for Tables VII-8 to VII-11 (10% measurement errors).....	91
VII- 8. Market model - 10% measurement noise: OLS estimates of parameters in equation 1.....	93
VII- 9. Market model - 10% measurement noise: OLS estimates of parameters in equation 2.....	94
VII-10. Market model - 10% measurement noise: OLS estimates of parameters in equation 3.....	95
VII-11. Market model - 10% measurement noise: OLS estimates	

of parameters in equation 4.....	96
X- 1. One-dimensional model: values of variables for Figures X-1 and X-2.....	151
X- 2. One-dimensional model: values of variables for Figures X-3 and X-4.....	154
X- 3. One-dimensional model: manual minima for the Likelihood function.....	156
X- 4. One-dimensional model: manual minima for the Likelihood function. (10% equation noise).....	159
X- 5. One-dimensional model: Estimates of A, Q and R for 1% and 10% equation noise.....	170
X- 6. Two-dimensional Oscillator: OLS estimates.....	173
X- 7. Two-dimensional Oscillator: LIKALM estimates - Convergence zone under ideal conditions.....	177
X- 8. Two-dimensional Oscillator: LIKALM estimates - Wrong initial values for X _{Eo} and S _{Eo}	180
X- 9. Two-dimensional Oscillator: LIKALM estimates - Estimation of A, Q and R.....	182
X-10. Two-dimensional Oscillator: LIKALM estimates - Wrong values for Q and R.....	184
X-11. Two-dimensional Oscillator: LIKALM estimates - Different combinations of Q and R based on the Least Squares Residual S.....	186
X-12. Two-dimensional Oscillator: OLS and LIKALM estimates with different noise sequences.....	187
X-13. Three-dimensional model: OLS estimates	192

X-14. Three-dimensional model: LIKALM estimates	193
X-15. Noise sequences for Market model.....	197
X-16. Market model: OLS estimates with new form of equation 4.....	198
X-17. Market model: Scaled parameter values.....	200
X-18. Market model: OLS estimates with new form of equation 4 (Scaled data).....	202
X-19. Market model: LIKALM estimates of parameters in equation 1 only.	203
X-20. Market model: LIKALM estimates of parameters in equation 3 only. (Under different conditions)....	204
X-21. Market model: LIKALM estimates of parameters in equation 4 only.	205
X-22. Market model: LIKALM estimates of parameters in equations 1 to 3.	207
X-23. Market model: LIKALM estimation of entire model. ($A_0 = A - 10\%$)	208

List of Figures

Figure		Page
I- 1.	Four elements of structure in the System Dynamics methodology.....	7
I- 2.	A simple feedback loop.....	8
I- 3.	Flow diagram of part of the Urban Dynamics model..	10
II- 1.	The process of System Identification.....	23
IV- 1.	Engineering formulation of the Identification problem	49
V- 1.	Observationally equivalent models.....	61
VI- 1.	The experimental procedure.....	70
VI- 2.	Possible non-ideal conditions for testing an estimation technique.....	71
VII- 1.	Major feedback loops in the Market-Growth model..	74
VII- 2.	Behavior of the Market-Growth model.....	76
IX- 1.	Reinitialization at each point in time in OLS simulation.....	113
IX- 2.	Reinitialization at each point in time in LIKALM simulation.....	115
IX- 3.	Iterative scheme for the LIKALM algorithm.....	124

IX- 4.	Flow chart for the LIKALM algorithm.....	138
X- 1.	One-dimensional model: simulation of trajectory using OLS (first iteration).....	148
X- 2.	One-dimensional model: simulation of trajectory using OLS(last (second) iteration).....	149
X- 3.	One-dimensional model: simulation of trajectory using LIKALM (first iteration).....	152
X- 4.	One-dimensional model: simulation of trajectory using LIKALM (last (fourth) iteration).....	153
X- 5.	One-dimensional model: plots of $L(A)$, $L(Q)$ and $L(R)$ ($Q=1\%$, $R=10\%$).....	158
X- 6.	One-dimensional model: plots of $L(A)$, $L(Q)$ and $L(R)$ ($Q=10\%$, $R=10\%$).....	160
X- 7.	Two-dimensional model: trajectory with and without noise inputs.....	172
X- 8.	Two-dimensional model: sensitivity of model to change in the parameters.....	179
X- 9.	Three-dimensional model: noiseless trajectory....	190
X-10.	Sensitivity of Market-Growth model to parameter changes.....	211

NOTATION

b : parameter to be estimated

$\hat{b}$: estimate of b

$\hat{b}$: best estimate (if $\hat{b}$ has already been used as an estimate)

x : state

$\hat{x}$: estimate of state

z : measurement

$\hat{z}$: predicted measurement

A : matrix (all FORTRAN variables are also capitalized)

$\hat{A}$: estimate of A

A' : transpose of matrix A

A^{-1} : inverse of matrix A

$|A|$: determinant of A

$\text{tr}(A)$: trace of matrix A

$\underline{u}$: vector

$\underline{u}'$: transpose of a vector

u' : scalar, different from u

u : time dependency (also indicated as $u(t)$)

T : sample size

$Z(T)$: when defined, set of values of $z(t)$ from $t=0$ to $t=T$

$\bar{x}$: mean value of x

$E(x)$: expectation of x

$\text{Var}(x)$: variance of x

$\text{Cov}(x,y)$: covariance of x and y

$p(x)$: probability function of x

plim x: probability limit
s : variance of a scalar
S : covarince matrix of a vector
•: scalar, vector or matrix multiplication (used only when needed for more clarity)
*: multiplication in FORTRAN
L: Likelihood function
l: letter "l" (as opposed to number "1")
Log: Natural logarithm
d: partial differentiation operator

NOTES

1. Equations are numbered separately for each section. The section number appears before the equation number only when referring to an equation in another section.
2. The parameter vector, equation and measurement noise covariance matrices are noted A, Q and R, respectively, in the FORTRAN equations and in all experimental results.

I. INTRODUCTION

A. THE NEED OF OUR TIMES

We live in a time of social crisis. One of the major underlying causes for this crisis is that the systems we live in have grown so complex that no one can claim to understand fully the mechanisms involved in these systems or predict their reaction to a given policy change.

According to Warfield [1], much of the difficulty in coping with complex systems comes about because the rate of change in society is faster than the rate at which we can respond effectively.

And societal problems are highly complex:

"They involve multiple interactions and feedback among various elements of society; they are highly interdisciplinary and transdisciplinary, with social, economic, political and emotional factors intertwined with more quantifiable factors of physical technology; they cannot be shown easily to be "solved" by experiment; and the implementation of the presumed amelioration to a problem changes the system in complex ways..." [1]

It was crucial therefore, that "new, improved and systematic ways of coping with such complex problems be developed"...

Perhaps one of the most promising applications of technology available to decision makers to help them deal with

increasingly complex societal problems could be broadly labelled "Systems Analysis", the first truly interdisciplinary approach, stemming out from the post-war field of "General Systems Theory" [2], and the closely related fields of "Cybernetics" and "Information Theory" (see for example [3] and [4]).

The growing family of world models associated with Forrester [5], Meadows *et al.* [6], Mesarovic and Pestel [7] (to quote only a few) have demonstrated the power of systems concepts and modeling for organizing one's thought and for tackling seemingly intractable problems.

B. TWO APPROACHES TO MODELING

There are two main trends in model-building, based on entirely different approaches: "Econometrics" and "System dynamics".

These two types of models differ both in their construction and in their purpose.

A.1 Econometric models

Econometric models, derived from the methods of the social sciences, rely heavily on formal statistical techniques and this tends to influence their structure. Because the available estimation techniques are mostly linear, econometricians will tend to linearize any nonlinear relationship, to omit all variables which cannot be measured or for which data is not available and all feedback relationships which cannot be derived from statistical data.

The regression procedures will also often reject any hypothesized relationship that has not played a significant role in the past (though it may be crucial in the future). The usual explanatory equation expresses each endogenous variable as a function of its assumed direct causes :

$$y_i = f_i(y_1, y_2, \dots, y_{i-1}, y_{i+1}, \dots, y_G, x_1, \dots, x_K) \quad (1)$$

where x_j are predetermined variables (exogenous and lagged endogenous).¹ Thus, an econometric model is a system of G equations explaining the G endogenous variables.

Specification of the structure of an econometric model is the most crucial step in that later statistical estimation and validation procedures assume that the structure is correct.

Brown [8] defines specification as the process of deciding on the hypothetical structure of a model, preparatory to testing this with observed data, and finally estimating its parameters. Clearly then, the first major part of specification is to decide which variables belong to which side of the equations. The next phase of specification involves deciding on the form of each equation (very often chosen as linear).

" Good specification is vital for econometric model-building. Without careful attention to this phase of the work, we can easily settle for spurious

¹ For definition of terms see page 27

relations, selected by searching only for good fits and high correlations. With small samples of data, it is easy to hammer out hypotheses to conform closely to the sample. Specification is then biased toward the particular sample..." [8]

However, the major decision criteria come directly from economic theory...

" The way in which careful specification can prevent acceptance of faulty structures is to let economic theory dominate over the statistical tests, when these are being made from small samples. Thus, when the statistical tests favour an apparently inferior hypothesis, the econometrist should put more weight on what he believes to be superior theoretical specification, while checking his preferred theory for possible flaws. Decisions and judgements of this kind of course require a considerable experience and expertise in the workings of the macro-economy..." [8]

Thus, although econometricians claim their models are based on measured data, the judgemental process in the choice of an equation predominates. Because econometric models do not attempt to explain the internal mechanisms that generate the observed data, they typically contain many exogenous variables (variables that are not determined by the model).

Many large-scale econometric models of national and

provincial economies have been constructed in the last two decades and are mostly used by government agencies and central banks in short-term forecasting and policy simulation. Although they perform well in short-term forecasting, they do not generally pick up major turning points, because they assume the continuation of present trends, do not attempt to explain mechanisms and are tested only with mild changes in policy variables.

A.2 System dynamics models

The structure of system dynamics models, on the other hand, is not limited by the difficulties inherent in formal statistical estimation techniques. System dynamics models rely on a philosophy of structure in systems that results in many factors that make formal estimation very difficult, e.g., strong nonlinearities, multiple feedback loops, unmeasured variables, delays, etc...

Because of its relative novelty, this approach will be developed in somewhat more detail here than the econometric one.

Most of the following description of system dynamic's methodology is based on "Urban dynamics" [9], but all Forrester's models basically have the same features.

" In the 1960's, Professor Jay Forrester of Massachusetts Institute of Technology pulled together some existing techniques of feedback control, mixed in some of Norbert Wiener's principles of cybernetics, added inventions of his

own, and developed a methodology that is known as System Dynamics. " (McLeod,[10])

Structure in a system is seen as having four significant hierarchies, as illustrated in Figure I-1 (from Sage [11]).

1. The closed *boundary*
2. The *feedback loops* as the basic structural elements within the boundary
3. *Level (state) variables* representing accumulations within the feedback loops, and *Rate (flow) variables* representing activity within the feedback loops
4. *Goal, observed condition, detection of discrepancy, action* based on discrepancy as components of a rate variable.

The boundary of a system is chosen to include those interacting components *necessary to generate the modes of behavior of interest*: this means that the latter are not imposed from the outside but created within the boundary. The system can however be influenced by external forces, but these should be random, not give the system its intrinsic growth and stability characteristics.

The dynamic behavior of the system is generated within the feedback loops, which are the fundamental building blocks of systems. Figure I-2 shows the simplest possible feedback-loop structure.

A feedback loop is a structure within which a decision point - the rate equation (symbolized by a valve) - controls

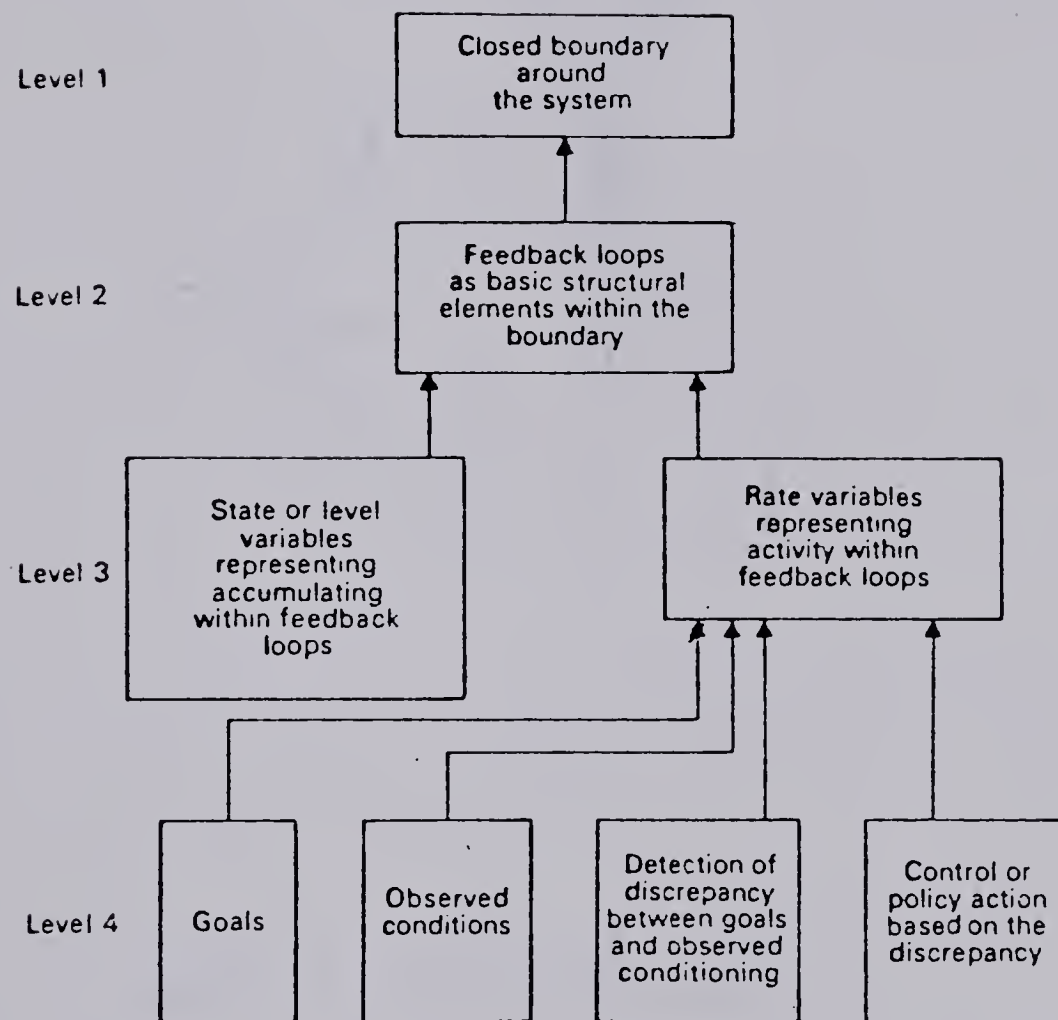


FIGURE I-1: Four elements of structure in the System Dynamics methodology.

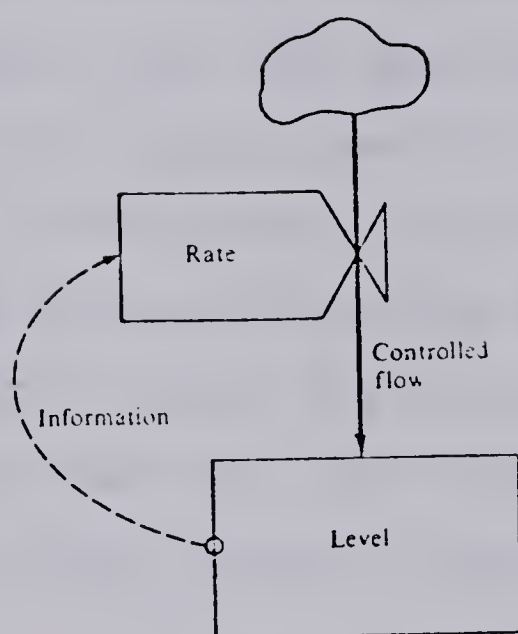


FIGURE I-2: A simple feedback loop.

the flow or action stream. The action is accumulated (integrated) to generate a system level. Information about the level is the basis on which the flow rate is controlled. The "cloud" describes an unlimited quantity.

The rate equations are the statements of system policy. They determine how the available information is converted to an action stream. Within each rate equation there is implicitly or explicitly a goal toward or away from which that decision point in the system is striving. There is also a process whereby the observed condition of a system is detected. A rate equation states the discrepancy between the goal and the observed condition. And finally, the rate equation states the action that will result from the discrepancy.

Once all the rate and level variables have been identified (into three major sectors: population, industry and housing, in the case of Urban Dynamics, for example) and carefully defined, once the causal relationships are determined, the various building blocks are interconnected using a causal loop diagram to form a flow diagram, a part of which is illustrated in Figure I-3.

The rate equations are postulated and the parameters specified and then all the equations are written in a DYNAMO language, typically using a set of "normal" conditions which are modified by multipliers, representing the deviation of the actual system from the "normal", as illustrated below.

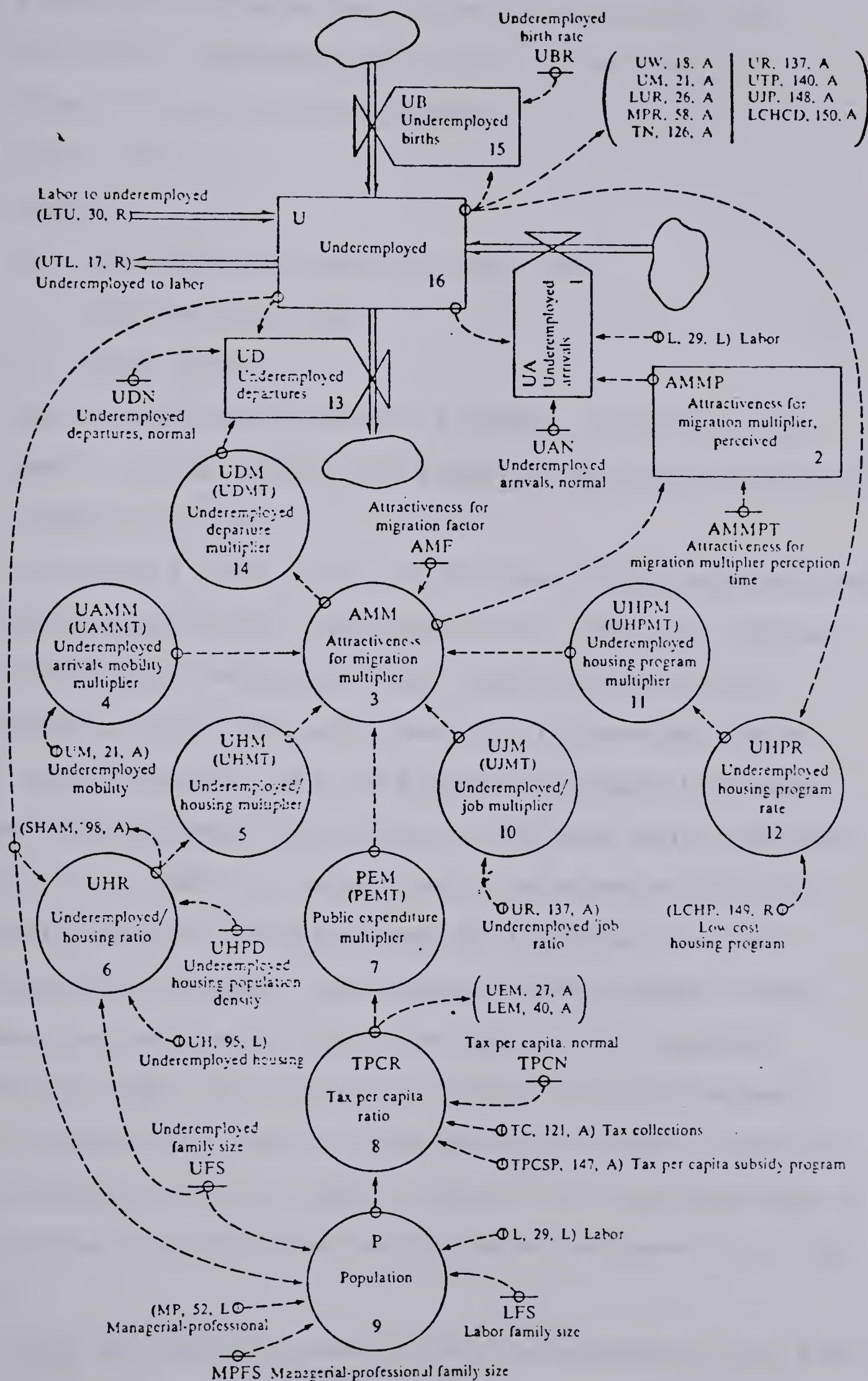


FIGURE I-3: Flow diagram of part of the Urban Dynamics model (from [9])

A typical rate equation is the one describing the arrival of underemployed into the urban area :

$$UA.KL = (U.K + L.K)(UAN)(AMMP.K)$$

$$UAN = .05$$

where

UA = UNDEREMPLOYED ARRIVALS (MEN/YEAR)

U = UNDEREMPLOYED (MEN)

L = LABOR (MEN)

UAN = UNDEREMPLOYED ARRIVALS, NORMAL (FRACTION/YEAR)

AMMP = ATTRACTIVENESS-FOR-MIGRATION MULTIPLIER PERCEIVED (DIMENSIONLESS)

The first term, $(U+L)$, is the sum of the underemployed and labor populations. The second term, UAN, is a "normal" multiplier which represents the fraction of existing inhabitants $(U+L)$ that would come to the area each year. With a UAN value of .05, the statement is made that under normal circumstances the inflow to the area would represent 5% of $(U+L)$. AMMP is a modulating term representing the attractiveness of the area compared to normal. As the attractiveness changes, the value of AMMP changes taking values less or greater than 1. K and $L = K+1$ represent points in time. The rate UA is constant during the small unit interval KL, and its value during this period depends on the values of U, L, AMMP at time K. All the equations of the system are thus iteratively updated and generate a time path.

Once the whole program is thus formulated and verified,

the model is ready to run. A reference is chosen and then various changes in equations and parameters are made to test the sensitivity of the model. Changes in policy variables are also performed, each new run being compared to the reference run in order to analyze the effects of that given change.

Typically, the models are highly *insensitive to most of the parameter changes*; "common" or "traditional" policies designed to solve the system's problems are found to be ineffective or making matters worse; a few, unexpected, couterintuitive policies are shown to be more likely to deeply alter the system's behavior in the desired way...

Similar patterns were followed in "Industrial Dynamics" [12] and in "World Dynamics" [5] which led to the well-known "Limits to growth" [6].

All of these models caused a major break-through in traditional ways of thinking and have been highly praised by the scientific and decision-making communities. All of them have also been criticized because of the many over-simplifications made (see for example Sage [11], Ansoff and Slevin [13], and the "Sussex Group" [14]).

It appears from Forrester's writing that "*expert opinion*" rather than real data were often used in constructing the many nonlinear relationships and in evaluating the multipliers used. He has been criticized for developing a model with built-in bias, for not explicitly stating a value system, and for making untested and possibly

invalid assumptions...

Although some of the critiques are justified - no model is perfect, specially at this early stage of the technology - they have, in general, not been constructive, reflecting conservationist attitudes, as it appears from the responses to these critiques (see Forrester [15], Meadows *et al.* [14]

To summarize, in Forrester's words [16]:

"Compared to an econometric model, a system-dynamic model:

1. Makes greater use of *descriptive information* and managerial and political experience,
2. Incorporates a broader range of variables and encompasses the many relevant disciplines outside of economics,
3. *Uses numerical data* from real life in a different way - in model construction *to complement descriptive information*, in validation to compare with the corresponding output data from the model,
4. Generates social and economic fluctuations and growth from the internal feedback structure without using exogenous variables to cause the change to occur,
5. Includes important social and psychological variables for which *statistical data are not available*,
6. Explains how structural change occurs as the economy moves into new modes of behavior that are not represented in past time-series data,
7. Facilitates incorporating the wide range of *nonlinear*

structures that generate so much of observed real-world behavior,

8. Emphasizes the conservation of flows by including the buffer stocks that decouple the instantaneous flow rates,
9. Distinguishes more sharply between real variables, their money value, and information about them to capture the dynamic interactions between these elements of a real-world system,
10. Encourages construction of a deeper substructure of *feedback loops* to represent causal mechanisms underlying macroeconomic behavior,
11. Organizes structure so that *each parameter has independent real-life meaning in the operating world and can be individually drawn from and checked against descriptive and quantitative information available at the place in the real system to which the parameter applies,*
12. Serves as a more effective communication medium for resolving disagreements because of the way both model structure and parameters correspond with descriptive knowledge in the operating world,
13. Places more emphasis on the importance of internal structure,
14. Focusses more on understanding the reasons for observed behavior and on developing policies to produce better behavior, and *focusses less on prediction,*

15. Combines over a greater time span short-term with long-range human objectives,
16. Permits a wider diversity of contact between the model and the real world to make *validation* more persuasive..."

This lengthy presentation of the main features of system dynamics models gives an idea of the theory on which they are built, namely that "the real social systems belong to the class of nonlinear integration-feedback systems" , theory which according to Senge [17], is a necessary foundation for any analysis of the usefulness of a statistical technique for modeling practice .

"The generality of an analysis of estimator accuracy depends on whether or not the data-generating model in the experimental laboratory (or the assumed "true system" in the theoretical analysis) belongs to the same class of systems as real-life social systems. If the data-generating model does not belong to the proper class of systems, the investigator has little basis for inferring from the laboratory analysis how a statistical technique will perform in real-life applications." [17]

C. SENGE'S ATTEMPT

In a recent paper, [18], Peter SENGE has attempted to clarify issues underlying the controversy over statistical parameter estimation in feedback models.

Experiments with Ordinary Least Squares (OLS) and Generalized Least Squares (GLS) estimation, using Forrester's 'Market growth' model [19], show that the parameter estimates derived from these methods are highly sensitive to errors in data measurement, especially when a model's feedback structure is not completely known. In doing so, he has developed an experimental procedure to evaluate the estimation techniques, and a range of experimental conditions exploring the effect on estimation results of a particular facet of real life model-building or data measurement.

This framework provides useful guidelines for conducting further investigation of these issues.

He concludes by suggesting that the "ability of alternative estimation techniques to improve upon the performance of OLS and GLS should be explored. Statistical methods which warrant investigation include alternative econometric techniques, as well as techniques developed in such fields as engineering..." [18], and also that the accuracy of these techniques, including OLS and GLS, should be examined for alternative feedback models.

D. THIS THESIS

It is in the light of all the above facts that the present investigation proposes to try alternative econometric and engineering methods to estimate the market growth model.

As a first step, sections II, III and IV attempt to review the main trends in system identification, both in the econometric and the engineering literatures. Although there are many similarities between the two fields, they have developed in different directions, used different terminologies, and emphasized different types of models. Current research seems to be moving in the direction of bridging the gap between the two disciplines, particularly in the attempts to use Kalman filtering concepts in econometrics, but by and large, current research areas still differ quite largely.

The effect of noise on estimation results are given particular attention.

The most useful econometric method, in our case, is found to be the LISREL model². However, although it allows for Maximum Likelihood estimation when both equation and measurement errors are present, this model does not accomodate dynamic or nonlinear systems.

The most promising estimation algorithm is found in the engineering literature and is known as the "Filtering form of the Maximum Likelihood algorithm". The detailed explanation of this approach is delayed until section IX.

The identifiability problem, which is the cause of much trouble in multivariate system estimation, is presented briefly in section V. Although this is a main topic of research in both fields, no practical way of checking for

²To be explained in section (III-D)

identifiability, other than an analysis of singularity problems is usually provided...

In a second part, section VI describes the experimental framework; section VII is an attempt at reproducing Senge's experiment in order to ascertain that the results obtained are comparable. Section VIII describes the LISREL model which was tried on experimental smaller order dynamic models and then on the Market model; but the results were disappointing and are not presented here. In section IX, the theory of the Filtering form of the Maximum Likelihood algorithm is presented, along with a simple one dimensional example. The equations for the general case are also derived here.

Section X describes the main emphasis of this thesis: the development of a computer program for the estimation of parameters using the Filtering form of the Maximum Likelihood algorithm and the experimental results obtained with this program on four models of varying complexity. The performance of this algorithm is found to be very acceptable for small models, specially in the light of the OLS results, despite the problems in choosing values for, or estimating the various noise covariance matrices. It does however present many problems if used on a larger and unstable model such as the Market-Growth model.

Finally, section XI presents the conclusions for this set of experiments, examines the applicability of such a program for "Real-World" problems, and suggests an

alternative approach for model validation.

II. SYSTEM IDENTIFICATION: AN OVERVIEW

The use of mathematical models to describe the behavior of a physical phenomenon is now common practice in a great number of scientific disciplines. More recently however, the systems under study have drastically grown in complexity and have also extended to social and economic disciplines. The need for control engineers to have accurate knowledge of their plants and for socio-economic model-builders to present reliable models of systems to decision-makers has caused an increasing interest in the techniques of model-building.

If all the phenomena under study were known with certitude, and if exact knowledge of these systems was possible, exact calculation of model parameters would also be possible and we would have deterministic models.

Unfortunately, there are usually many unknown factors (lack of exact knowledge of the system, or even if this is possible, there is usually measurement noise or disturbances, etc...). Being incapable of writing exact equations or to predict behavior with certitude, we have to refer to probabilistic methods. Based on observations, we have to build physical insight into the problem being studied and come up with with a *model*, defined in Eykhoff [20] as:

" a representation of the essential aspects of an existing system (or a system to be constructed) which presents knowledge of that system in a *usable*

form".

The revolutionary work of Kalman [21] and others has led to modern control theory in which the models used are with a few exceptions parametric models in terms of state equations. The desire to determine such models from experimental data has renewed the interest of control engineers in parameter estimation and related techniques.

Independently, econometricians have developed a number of tools for their econometric model-building (mainly regression techniques) and created a wealth of statistical techniques for the estimation and testing of their models (see Theil [22], for example).

Our knowledge about systems is often in terms of time-series, i.e., a set of successive values for $t = 1 \dots T$ for the input $u(t)$ and the measured output $z(t)$.

Some "a priori" knowledge about the system and even some approximate knowledge of the values of the parameters in the model are available in some cases. Our main concern here, given $u(t)$ and $z(t)$ on a period of time, is to determine the (differential) equations that describe the system's behavior.

A model represents three types of knowledge:

1. Structure
2. Parameters' values
3. Values of dependent variables.

The process of "System Identification"³ is then an iterative process to determine these three types of knowledge since they are highly interdependent, and has been illustrated by Box and Jenkins [23] as follows:

1. From the interaction of theory and practice, a successful class of models for the purposes at hand is considered.
2. Because this class is too extensive to be conveniently fitted directly to the data, rough methods for identifying subclasses of these models are developed. These employ data and knowledge of the system to suggest an appropriate class of "parsimonious" models which may be tentatively entertained.
3. The tentatively entertained model is fitted to data and its parameters *Estimated*
4. *Diagnostic* checks are applied with the object of uncovering possible lack of fit and diagnosing the cause.

The process is illustrated in Fig II-1.

Although this is intended for a "Box-Jenkins" type of statistical identification, it gives an idea of the general identification process.

The field of system identification and parameter estimation is rather complex. This complexity is due to the variety of applications, in fields as different as

³ The term "identification" is used here in its engineering sense. The econometric equivalent is "specification and estimation", the term "identification" being used exclusively in the context of the identifiability problem.

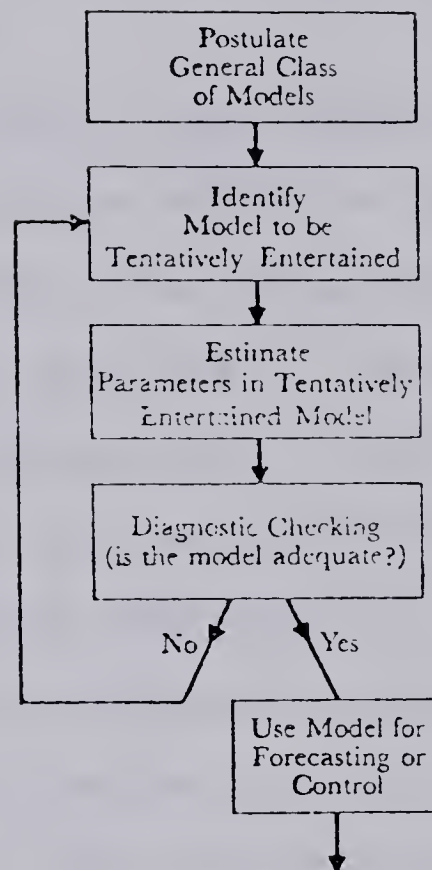


FIGURE II-1: The process of System Identification (from[23])

communications, power, mechanical, aeronautical and chemical engineering, physics, geology, economics, biology and medicine; to the variety of goals, whether of research, design or control; to the variety of conditions: what kinds of signals are available, what "a priori" knowledge about the system there is, etc...

The models can be single input-single output (SISO), multiple input-multiple output (MIMO) or combinations thereof, linear or nonlinear (in the parameters or in the variables), continuous or discrete,... An excellent survey of all these models and approaches is given by Eykhoff [20]. Each has been extensively studied, and different approaches and algorithms developed for each specific case.

The two forms which are general enough to include the major part of the identification practice are the engineering "*state-space*" model and the econometric "*simultaneous equations*" model.

The econometric and engineering practices of system identification have developed in parallel for a long period of time. Although in general they use well known statistical techniques, differences in the vocabulary, notation, models considered, availability of data, purposes of identification and the emphasis of research made cross-communication very difficult or non-existent. It is only in the past few years that some attempts have been made at bridging this gap. In this respect, Mehra's outstanding article in *Annals of Economic and Social Measurement* [24] is worth mentioning.

To make communication between researchers in the two fields easier, he proposes the state-space model of control theory as a unifying link and discusses the relationship of this model to the "simultaneous equation" model as well as the use of process and measurement noise in the econometric context. He also develops an algorithm for model structure determination and estimation of parameters. Mathematical economists have manifested an increasing interest in control theory and its applications to econometrics (see [25], [26] for example). In particular, a number of articles examine problems of parameter estimation for linear and nonlinear econometric models from the point of view of Kalman filtering. ([24], [27], [28], [29]).

The next two sections attempt to review briefly the main trends in both fields as well as the emphasis of current research.

III. SOME ECONOMETRIC ESTIMATION TECHNIQUES

In his experiment, Senge [17] considers only OLS and GLS which are estimation techniques derived for the standard single-equation linear regression model and for the combination of several linear relations, respectively.

However, most econometric models attempt to model the general interdependence of economic phenomena and lead to systems of simultaneous equations where, in general, each equation has several dependent variables which also occur in other equations. A number of methods have been developed to deal with this situation (see for example [22],[30],[31]), some of which will be presented here.

A. NOTATIONS AND ASSUMPTIONS

The basic "simultaneous equations" econometric model is a system of G equations in G endogenous variables, K predetermined variables, and S parameters, and can be written in implicit form as:

$$F_j(y_1, \dots, y_G, x_1, \dots, x_K, a_1, \dots, a_S) = 0 \quad j=1 \dots G \quad (1)$$

where:

y_i $i=1 \dots G$ are current *endogenous* variables: they are determined by the system.

x_k $k=1 \dots K$ are *predetermined* variables. they contain:

- current and lagged exogenous variables
- lagged endogenous variables.

a_l $l=1 \dots S$ is a set of *parameters*

If the system is assumed to be linear (which is the case of most econometric models), equation (1) becomes:

$$\sum_{i=1}^G b_{ji} \cdot y_i + \sum_{i=1}^K c_{ji} \cdot x_i = 0 \quad j=1 \dots G \quad (2)$$

If we have T observations on this system, introducing the observation index $t = 1 \dots T$ and the disturbance terms⁴ u_t , the system for observation t is:

$$\sum_{i=1}^G b_{ji} \cdot y_{it} + \sum_{i=1}^K c_{ji} \cdot x_{it} = u_{jt} \quad j=1 \dots G \quad (3)$$

or, in matrix notation:

$$B \cdot \underline{y}_t + C \cdot \underline{x}_t = \underline{u}_t \quad t=1 \dots T \quad (4)$$

where:

B is $(G \times G)$ nonsingular

C is $(G \times K)$

$\underline{y}_t$ is $(K \times 1)$, $\underline{x}_t$ is $(K \times 1)$ and $\underline{u}_t$ is $(G \times 1)$

Finally, the notation for the entire system of T observations is:

$$Y \cdot B' + X \cdot C' = U \quad (5)$$

where:

$Y = (\underline{y}_1 \dots \dots \underline{y}_T)'$ is $(T \times G)$:

matrix of observations on the jointly dependent variables,

$X = (\underline{x}_1 \dots \dots \underline{x}_T)'$ is $(T \times K)$ and is assumed to have rank

K:

⁴The term "structural disturbance" or "residual" is introduced in econometric equations to account for the neglected determining factors explaining the left-hand variable. The assumptions on these residuals are presented below

matrix of observations on the predetermined variables

$U = (\underline{u}_1, \dots, \underline{u}_T)'$ is $(T \times G)$:

matrix of structural disturbances

The basic assumptions on the disturbance terms are:

1. Zero mean: $E(\underline{u}_t) = 0 \quad t=1 \dots T$
2. Unknown but finite covariance matrix: $E(\underline{u}_t, \underline{u}_t') = \Sigma$
3. No autocorrelation: $E(\underline{u}_t, \underline{u}_{t+s}') = 0$ if $s \neq 0$
4. No correlation with predetermined variables:

$$E(\underline{u}_t, \underline{x}_t') = 0$$

B. SOME ESTIMATION TECHNIQUES

B.1 Desirable properties for an estimator

Theoretically, our knowledge about the parameter of a model, b , whose estimate is $\hat{b}$ could be expressed in terms of the probability density function $p(\hat{b}, T)$ if T is the number of data points available. In practice however, this function is reduced to its most significant characteristics:

- "expected value" : $E(\hat{b})$
- "bias" : $E(\hat{b}) - b$
- "covariance" : $Cov(\hat{b}) = E\{(\hat{b} - E(\hat{b})) \cdot (\hat{b} - E(\hat{b}))'\}$

Some desirable properties of estimators are then:

- "Unbiasedness": if for each t : $E(\hat{b}) = b$
- "Consistency": if $\lim_{t \rightarrow \infty} P(|\hat{b} - b| < \epsilon) = 1$ or $\text{plim } \hat{b} = b$
- "Efficiency": if for all unbiased estimates $\hat{c}$,

$$Cov(\hat{b}) \leq Cov(\hat{c})$$

$$\text{or } \det\{Cov(\hat{c}) - Cov(\hat{b})\} \geq 0$$

i.e. b has "minimum variance".

If the first and third properties hold only for $t \rightarrow \infty$, they are "asymptotic".

Different methods of estimation of identified equations (for the "identifiability" problem, see section V) for simultaneous equation models have been proposed in the literature. These can be classified under the category of single equation methods (or Limited Information methods) and systems methods (or Full Information methods).

In single equation methods, we estimate each equation separately using only the information about the restrictions on the coefficients of that particular equation.

In systems methods, we estimate all equations jointly using the restrictions on the parameters of all equations as well as the covariances of the residuals.

B.2 Single equation methods

The most commonly used single equation methods are Ordinary Least Squares (OLS), Indirect Least Squares (ILS), Two-Stage Least Squares (2SLS), and Limited Information Maximum Likelihood (LIML).

For the purposes of single equation estimation, the equation for *fixed* $j = J \leq G$, in (3) can be written, after normalization, as:

$$y_{jt} = \sum_{i=1}^H b_{ji} \cdot y_{it} + \sum_{i=1}^L c_{ji} \cdot x_{it} + u_{jt} \quad (6)$$

where $y_1, \dots, y_H$ are the other endogenous variables included

in equation J,

$x_{1t} \dots x_{Lt}$ are the predetermined variables included in equation J,

and the b_{ji} and c_{ji} are corresponding coefficients⁵.

Introducing the observations for $t = 1 \dots T$, we get:

$$\underline{y}_J = Y_J \bullet \underline{b}_J + X_J \bullet \underline{c}_J + \underline{u}_J \quad (7)$$

where Y_J ($T \times H$) and X_J ($T \times L$) are respectively matrices of other endogenous and predetermined variables included in equation J, and $\underline{b}_J$ ($H \times 1$) and $\underline{c}_J$ ($L \times 1$) the corresponding coefficient vectors. $\underline{y}_J$ and $\underline{u}_J$ are ($T \times 1$).

Defining $Z_J = (Y_J \ X_J)$ and $\underline{a}_J = (\underline{b}'_J \ \underline{c}'_J)'$, equation (7) becomes:

$$\underline{y}_J = Z_J \bullet \underline{a}_J + \underline{u}_J \quad (8)$$

Ordinary Least Squares (OLS)

The OLS estimate of the parameter set $\underline{a}$ is given by minimizing the criterion:

$$S = (\underline{y} - Z\underline{a})' \bullet (\underline{y} - Z\underline{a}) \quad (9)$$

Setting $dS/d\underline{a} = 0$ yields:

$$(Z'Z) \bullet \underline{a} = Z' \bullet \underline{y}$$

⁵ They, as well as u , are different from those used in equation (3). They are now of a different sign and divided by b for normalization

or in our case:

$$\underline{a}_{JOLS} = (Z_J' Z_J)^{-1} \bullet Z_J' y_J \quad (10)$$

This estimate is biased and does not have the large sample property of consistency because of the correlation of the residuals with the y_J , which violates one of the basic assumptions of OLS⁶.

Consider the standard linear model:

$$y = X \bullet \underline{b} + \underline{u} \quad (11)$$

X , the $(T \times K)$ matrix of explanatory variables, is for this model assumed independent from $\underline{u}$ $(T \times 1)$ vector. Then:

$$\underline{b} = (X' X)^{-1} \bullet X' y = \underline{b} + (X' X)^{-1} \bullet X' \underline{u} \quad (12)$$

so that:

$$E(\underline{b}) = \underline{b} \quad \text{since } E(X' \underline{u}) = 0.$$

In the case of equation (10) however, since the Z_J terms contain the matrix y_J of endogenous variables that are not independent of the residuals, the estimate $\underline{a}_J$ will be biased.

Nonetheless, OLS is still used because it was found to be more robust against specification errors than some other methods.

⁶ The OLS assumptions are that the residuals satisfy :

1. Zero mean $E(u_t) = 0$ for all t .
2. Homoscedasticity or same variance $V(u_t) = s^2$ for all t .
3. Serial independence : u_t and $u_{t'}$ are independent for all $t \neq t'$
4. u_t are normally distributed.
5. x_t are independent of $u_{t'}$ for all t and t' . $E(u_t x_{t'}) = 0$ for all t and t' .
6. The elements of Z are linearly independent. Hence $(Z' Z)^{-1}$ exists.

Indirect Least Squares (ILS)

In ILS estimation, the reduced form equations (obtained from equation (5)),

$$Y = X \bullet P' + V \quad (13)$$

$$\text{where } P' = -C' \bullet (B')^{-1} \quad (K \times G)$$

$$\text{and } V = U \bullet (B')^{-1} \quad (T \times G)$$

are estimated by OLS, and lead to the structural parameters. If the equation is overidentified (see section V), this leads to multiple solutions. ILS is therefore not much discussed.

Two-Stage Least Squares (2SLS)

In 2SLS, we first estimate the reduced form equations:

$$Y_J = X \bullet P'_J + V_J \quad (14)$$

where P'_J ($K \times H$) and V_J ($T \times H$) are derived from equation (13) and the definition of Y_J (equ. (7)) by a suitable partitioning, to obtain the estimate of P'_J , denoted $\hat{P}'_J$. Let us now define:

$$\hat{Y}_J = X \bullet \hat{P}'_J \quad (15)$$

These values are not correlated with the disturbance terms. Using a basic Least Squares relationship we can write:

$$Y_J = \hat{Y}_J + V_J \quad (16)$$

where V_J is the ($T \times H$) matrix of estimated residuals.

The idea then is to replace Y_J in equation (7) by $\hat{Y}_J$ thus yielding:

$$\underline{y}_J = \hat{Y}_J \bullet \underline{b}_J + X_J \bullet \underline{c}_J + \underline{u}_J + \hat{V}_J \bullet \underline{b}_J \quad (17)$$

Now defining $\tilde{Z}_J = (\hat{Y}_J \ X_J)$ and $\underline{w}_J = \underline{u}_J + \hat{V}_J \bullet \underline{b}_J$, we obtain a new form of our equation:

$$\underline{y}_J = \tilde{Z}_J \bullet \underline{a}_J + \underline{w}_J \quad (18)$$

The new Least Squares estimate is now defined by:

$$\underline{a}_{J_{2SLS}} = (\tilde{Z}_J' \bullet \tilde{Z}_J)^{-1} \bullet \tilde{Z}_J' \bullet \underline{y}_J \quad (19)$$

This estimate is consistent, has the asymptotic property of efficiency and is the most used in econometric practice.

Limited Information Maximum Likelihood (LIML)

LIML is obtained by maximizing a rather complex and nonlinear Likelihood function and is shown (under the assumption of normality of the disturbance terms) to be consistent and asymptotically as efficient as 2SLS. Because of the computational burden 2SLS is usually preferred to LIML in econometric practice.

The method is briefly outlined in section (IV-B.2) in the context of the engineering literature and developed in more detail in section IX.

B.3 Systems methods

The two principal systems methods, Full-Information Maximum Likelihood (FIML), and Three-Stage Least Squares (3SLS) are generalizations of LIML and 2SLS and are shown to be consistent and equally efficient asymptotically.

Because they use more information, at each stage than the single equation methods they are asymptotically more efficient than the latter. Just as LIML, FIML entails computational problems.

NOTE

In addition to these methods, one can also consider what is known as the Instrumental Variables (IV) method, where a set of instrumental variables satisfying certain conditions⁷ [32] is selected and the estimator computed as:

$$\underline{a}_{IV} = (W' \bullet Z)^{-1} W' \bullet y \quad (20)$$

where W is a set of instruments,

for a single equation, or as:

$$\underline{a}_{IV} = (W \bullet Z)^{-1} W \bullet y \quad (21)$$

where W is a set of instruments, for the whole system.

It can be shown, [32], that ILS, 2SLS, 3SLS are special cases of IV estimation. Necessary and sufficient conditions for an IV estimator to be efficient are also derived in [32].

C. DEPARTURES AND CURRENT RESEARCH AREAS

Departures from the assumptions of the linear simultaneous equations model and of the standard linear model are treated extensively in the econometric literature and constitute the bulk of current research.

⁷In particular that they must not be correlated with the disturbance terms

Some of the topics of interest are presented below.

C.1 Multicollinearity

The OLS estimate does not exist when, contrary to the assumptions, the matrix X of explanatory variables⁸ has rank less than K , that is, if some of the explanatory variables are linearly dependent or nearly dependent so that it becomes difficult to disentangle their separate effects on the dependent variable. (This latter case results in large standard errors for the coefficients)

The problem with multicollinearity is not one of existence or not, but of how serious the problem is. There are only some rough rules of thumb by which we can decide whether multicollinearity is serious or not. (They mainly involve comparisons of the multiple correlation coefficients of the explanatory variables, the coefficient of multiple determination⁹ R^2 and the partial r^2)

Some suggested solutions are : dropping variables, using extraneous estimates, using ratios or first differences, principal components, etc... (see [33],[22] for more detail)

C.2 Autocorrelation

When the residuals are correlated over time in time-series it is often a sign of omitted variables (which are themselves serially correlated). It is then usually assumed that the residuals form an autoregressive process (AR), a moving average process (MA) or a mixed

⁸ e.g. in the case of the standard linear model: $\underline{y} = X \bullet \underline{b} + \underline{u}$

⁹ See p. 84

autoregressive moving-average process (ARMA) and they are dealt with as such.

The problem of autocorrelation is one of the most studied (see for example [34], [35], [36], [37], [38]) but we shall limit ourselves here to the examination of the effects of measurement noise, bearing in mind that autocorrelation does affect the estimates¹⁰.

C.3 Time-varying parameters

The recent interest of econometricians in the applications of the Kalman Filter¹¹ has contributed to a great increase in articles dealing with time-varying parameters, and structural changes in the course of estimation. This field is one of the first examples of cross-fertilization between the two disciplines. Some papers of interest are [24], [27], [28], [29].

C.4 Errors in the variables

This topic will be dealt with in more detail here since errors in the variables have such a drastic effect on estimation as reported by Senge [17] and therefore constitute our main subject of investigation.

The standard linear model in econometrics assumes that the neglected determining factors are the only cause of the fact that the explanatory variables do not fully account for the behavior of the left-hand variable. This is also an assumption of all the extensions of this model, in

¹⁰See Senge [17] for further elaboration on autocorrelation in the Market-Growth model

¹¹ See section (IX-E)

particular the simultaneous equations model.

The assumption is rarely justified since most econometric data contain observational errors and the variables that are actually measured are often proxies to what was really intended.

The problem

The effect of measurement noise on the OLS estimates is illustrated by the following example from [33].

Suppose that the model is :

$$y = bx + e \quad (22)$$

Instead of y and x , we measure :

$$\begin{aligned} y' &= y + v \\ x' &= x + u \end{aligned} \quad (23)$$

where u and v are measurement errors with zero means and variances s_u^2 and s_v^2 , respectively, and are both mutually uncorrelated and uncorrelated with the systematic parts.

That is :

$$E(u) = E(v) = 0$$

$$\text{Var}(u) = s_u^2$$

$$\text{Var}(v) = s_v^2$$

$$\text{Cov}(u,v) = \text{Cov}(u,x) = \text{Cov}(y,v) = \text{Cov}(x,v) = \text{Cov}(u,y) = 0$$

Equation (22) can be written in terms of the observed variables as:

$$y' - v = b(x' - u) + e$$

$$y' = bx' + w \quad (24)$$

$$\text{where:} \quad w = e + v - bu. \quad (25)$$

The reason we cannot apply OLS is that $\text{Cov}(w, x') \neq 0$. In

fact,

$$\text{Cov}(w, x') = \text{Cov}(-bu, x + u) = -bs_u^2 \quad (26)$$

Thus one of the basic assumptions of OLS is violated¹². If only y is measured with error and x is measured without error, there is no problem because $\text{Cov}(w, x') = 0$ in that case. Thus, given the specification (22) of the true relationship, it is errors in x that cause the problem.

If we estimate b by OLS applied to (24), we have :

$$b = \frac{\sum_{t=1}^T x' y'}{\sum_{t=1}^T x'^2} = \frac{\sum_{t=1}^T (x + u)(y + v)}{\sum_{t=1}^T (x + u)^2} \quad (27)$$

and:

$$\begin{aligned} \text{plim } b &= \text{Cov}(x, y) / (\text{Var}(x) + \text{Var}(u)) \\ &= s_{xy} / (s_x^2 + s_u^2) \end{aligned} \quad (28)$$

as all cross products vanish. Since $b = s_{xy} / s_x^2$, we have:

$$\text{plim } b = b / (1 + s_u^2 / s_x^2) \quad (29)$$

Thus, b will underestimate b . The degree of underestimation depends on s_u^2 / s_x^2 .

If there is more than one explanatory variable, and if all variables are measured with error, the expressions of the bias get very complicated, and both underestimation and overestimation are possible. [33]

Attempts to solve the problem

Theil [22] states that :

¹²See page 31

" Under the condition that the u 's and the v 's are normally distributed, one can apply the Maximum Likelihood method provided that the ratio s^2/s^2 of the error variances is known, but this provision is not usually fulfilled; and even if it is, the resulting estimators of b , s_u^2 and s_x^2 are not all three consistent..." [22, p. 609].

Other methods have also been attempted, none of which is really simple in application and some involve an "enlightened guess" of the error variance based on an evaluation of the quality of the data !...

A more recent approach [39] shows that econometricians have only recently started to examine the problem of measurement errors in Structural Equation Systems¹³

Here again, the misestimation caused by errors in the variables specially when autocorrelation or multicollinearity are also present, and the identification problems encountered [40], make the issue a very difficult one to resolve.

One of the most promising algorithms, the LISREL algorithm (for "Linear Structural Estimation by maximum

¹³In a structural equation model each equation represents a causal link rather than a mere empirical association. In a regression model, on the other hand, each equation represents the conditional mean of a dependent variable as a function of explanatory variables. It is this distinction that makes conventional regression analysis an inadequate tool for estimating structural equation models .[39]

Likelihood") is proposed by Joreskog [41], [42]. It is a Maximum Likelihood procedure based on the analysis of the covariance structures of structural equation models and allows for both errors in equations and errors in variables, but it is not Maximum Likelihood (only consistent) if used in systems with lagged endogenous variables.

A more detailed description of this method is given in section (III-D).

C.5 Small sample properties

All the properties established theoretically for estimators and described above are asymptotical or "large sample" properties. However for the practical choice of an estimation method in an econometric model, typically based on small samples of data (seldom exceeding 80 observations and frequently much smaller), the small sample properties of the various estimators are of crucial importance.

For reasons of mathematical intractability, relatively little is known about finite sample distributions of the various estimators. The exact finite sample distributions of Limited-Information Maximum Likelihood and Two-Stage Least Squares have been derived by Basmann and Nagar in certain special cases [43], [44], [45].

Nonetheless, some tentative results can be obtained in terms of Monte-Carlo methods (described here as in [31])

Monte Carlo methods are similar to the experimental procedure used by Senge (see section (VI-B.1)), except that

for a specified structure, repeated samples of disturbances are drawn by using appropriate tables of random numbers. The exogenous variables are combined with the disturbance values for each sample (usually of size 15 to 40) to generate values of the endogenous variables. The estimating methods under study are then applied to each sample set. This process is repeated a large number of times (typically 50, 100 or 200) and the resultant frequency distributions of estimates are studied in conjunction with the true values of the parameters in order to conjecture the small-sample properties of the estimators.

In studying individual parameters, three important criteria are usually distinguished. These are bias (B), variance (V), and mean square error (MSE) and are defined as:

$$B = \bar{b} - b \quad (30)$$

where : $\bar{b}$ is the mean of the sampling distribution of estimates and b is the true value of the parameter.

$$V = 1/N \sum_{i=1}^N (b_i - \bar{b})^2 \quad (31)$$

$$MSE = 1/N \sum_{i=1}^N (b_i - b)^2 = V + B^2 \quad (32)$$

The last formula shows that a biased estimate may show a smaller MSE than an unbiased one if it more than compensates for its bias by having a smaller variance.

A number of well-designed Monte-Carlo experiments are to be found in the econometric literature (see [46], [47],

[48], [49]), involving primarily LIML, 2SLS, OLS, and FIML. The general conclusions emerging from them are well summarized by Johnston in [31].

Some more recent results can be found in [50], [51], [52].

" Although these sampling experiments have yielded some results of broad usefulness, they have not been truly satisfactory, if for no other reason than the pervasive suspicion that the results are peculiar to the specification of the models and the structures used in the experiments ..." [53]

All these studies have assumed ideal conditions and when they have considered any departures, most of them have limited themselves to a single nonideal situation.

In view of the great disparity and absence of unity in the above studies, and of the nature of the models used in Monte-Carlo experiments, *no "a priori" insights can be drawn for the estimation of system dynamics feedback models...*

C.6 Estimation under Nonlinear Specification

Nonlinearity, in the econometric context described below, is a nonlinearity in the parameters and not in the variables, since nonlinearity in the variables, in general squares or logarithms and exponentials, is treated in econometrics as a linear problem.

Several extensions of the basic estimators such as

Nonlinear Ordinary Least Squares and Full Information Maximum Likelihood estimators [54], the Nonlinear Two-Stage Least Squares estimator [55], the Nonlinear Limited Information Maximum Likelihood and the Modified Nonlinear Two-Stage Least Squares estimators [56], and the Nonlinear Three-Stage Least Squares estimator [57] have been developed recently. (see also [58] , [59], [60], [61].)

They generally lead to estimators possessing all the desirable properties but they typically entail computational problems. Efficient computational methods for these estimators are very recent or still under investigation [62], [63].

D. THE LISREL MODEL

The following is a description of the LISREL computer program [64] developed by Joreskog and Sorbom and based on Joreskog's general method for analysis of covariance structures [41], [42].

The LISREL model allows for errors in both the equations and measurement errors, and yields estimates of the disturbance covariance matrix and the measurement error covariance matrix as well as estimates of the unknown coefficients in a structural equations system, provided that all these parameters are identifiable¹⁴.

If there are no errors of measurement, the method is

¹⁴ See section V for the identifiability problem

equivalent to the Full Information Maximum Likelihood (FIML) method¹⁵ provided that no constraints are imposed on the disturbance and independent variables covariance matrices.

The LISREL model consists of two parts: *the measurement model* and the *structural equation model* and is formalized as follows.

The true model is given by the set of structural equations:

$$B \bullet \underline{y} = C \bullet \underline{x} + \underline{u} \quad (33)$$

where,

B is (mxm) : coefficient matrix

C is (mxn) : coefficient matrix

$\underline{y}$ is (mx1) : vector of dependent variables

$\underline{x}$ is (nx1) : vector of independent variables

$\underline{u}$ is (mx1) : vector of random residuals

It is assumed that $\underline{x}$, $\underline{y}$ and $\underline{u}$ have zero mean, that $\underline{u}$ is uncorrelated with $\underline{x}$ and that B is nonsingular.

Instead of $\underline{x}$ and $\underline{y}$, we observe $\underline{X}$ and $\underline{Y}$ given by the measurement model:

$$\underline{Y} = H_y \bullet \underline{y} + \underline{v} \quad (34)$$

$$\underline{X} = H_x \bullet \underline{x} + \underline{w} \quad (35)$$

where :

- $\underline{Y}$ (px1) and $\underline{X}$ (qx1) are vectors of observed variables¹⁶,

- $\underline{v}$ and $\underline{w}$ are vectors of measurement errors in $\underline{y}$ and $\underline{x}$ respectively; they are assumed uncorrelated with the

¹⁵ See p. 33

¹⁶ $\underline{y}$ and $\underline{x}$ are then referred to as "latent variables".

latent variables,

- H_y ($p \times m$) and H_x ($q \times n$) are regression matrices of $\underline{y}$ and $\underline{x}$ on $\underline{y}$ and $\underline{x}$ respectively.

If in addition the following matrices are defined :

Φ ($n \times n$) : covariance matrix of $\underline{x}$

Ψ ($m \times m$) : covariance matrix of $\underline{u}$

Θ_v ($p \times p$) : covariance matrix of the noise $\underline{v}$

Θ_w ($q \times q$) : covariance matrix of the noise $\underline{w}$

it can be shown [41] that the covariance matrix Σ of $\underline{z} = (\underline{y}' \ \underline{x}')'$ is a function of H_y , H_x , B , C , Φ , Ψ , Θ_u and Θ_w .

The program is based on the analysis of the usual sample covariance matrix S :

$$S = 1/(N - 1) \sum_{i=1}^N \{(\underline{z}_i - \underline{z})(\underline{z}_i - \underline{z})'\} \quad (36)$$

The estimation problem is then essentially that of fitting the Σ imposed by the model to the sample covariance matrix S . This is done using the Likelihood function. The logarithm of the Likelihood function, omitting a function of the observations can be written as :

$$\text{Log } L = - (N - 1)/2 \{ \text{Log} |\Sigma| + \text{tr}(S \cdot \Sigma^{-1}) \} \quad (37)$$

This function is to be maximized with respect to $\underline{\theta}$, the vector of independent unknown parameters in B , C , Φ , Ψ , Θ_u , Θ_w .

The minimization is done by means of a modification of the iterative method of Fletcher and Powell [65] described by Gruvaeus and Joreskog [66]. The minimization method makes

use of the first derivatives and approximations to the second order derivatives of F^{17} and converges rapidly from an arbitrary starting point to a local minimum of F . If there are several minima of F , there is no guarantee that the method will converge to the absolute minimum. At the minimum of F , the Information matrix¹⁸ may be computed and used to compute the standard deviations for all estimated parameters.

The package also allows for a test of the goodness of fit of the model which, together with an inspection of the first derivatives of F can suggest improvements in the model specification (see [64] for more details).

The assumption of independent observations however, rules out dynamic models which include lagged endogenous variables observed with measurement errors. In that case, the LISREL method is consistent, but not Maximum Likelihood [41].

¹⁷ $F = 1/2 \{ \log |\Sigma| + \text{tr}(S \cdot \Sigma^{-1}) \}$ is the actually minimized function

¹⁸ See section (IV-B.2)

IV. THE ENGINEERING LITERATURE

The problem of identification and system parameter estimation has received extensive attention from the engineering community as well, specially in the last two decades, following the development of modern control theory and the pioneering work of Kalman.

In an excellent article, Astrom and Eykhoff [67] survey the state of the art citing over 230 references dealing with this problem. There is also an increasing amount of literature which deals with the problems of stochastic control in general (see for example [68], [69])

A. NOTATION

By and large, the engineering estimation techniques, Least Squares, Maximum Likelihood, Instrumental variables, etc., are very similar if not identical, to some of the econometric methods outlined in section III, although using different notations and terminology.

The following presentation of some main features of the engineering notation and methods is based on Eykhoff [20].

If the process output is given by:

$$\underline{z}(t) = F(\underline{u}, \underline{b}, \underline{n}) \quad (1)$$

and the output of a model by:

$$\underline{z}_m(t) = F(\underline{u}, \underline{b}, 0) \quad (2)$$

where:

$\underline{b}$ is the vector of process parameters

$\underline{u}$ is the process and model input

$\underline{n}$ is the noise

$\hat{\underline{b}}$ is the vector of parameter estimates,

the estimation problem may be formulated as a minimization of (see Eykhoff):

- the function(al) of a measurable quantity:

i.e.: $G\{\underline{z}(t, \underline{b}) - \underline{z}_m(t, \underline{b})\}$

- the function(al) of an unmeasurable quantity:

i.e.: $E\{G(\underline{b} - \hat{\underline{b}})\}$ E: expectation

In the control literature, a number of different estimation procedures have been developed. They differ predominantly in the criterion used for defining optimality and in the use of available "a priori" knowledge. Evidently, there is a close relationship between the amount of knowledge available and the type of "optimum" procedure that can be used to estimate the parameters.

The characteristics of the Least Squares, Markov or Generalized Least Squares, Maximum Likelihood and Bayes estimators will be considered here. The amount of assumed initial knowledge increases in this order. Starting from the Bayes estimator, one can derive the other estimators as a special case if there is less "a priori" knowledge available and under certain conditions of normality.

The problem can be stated as follows (see Figure IV-1):

We want to derive an estimator (relationship):

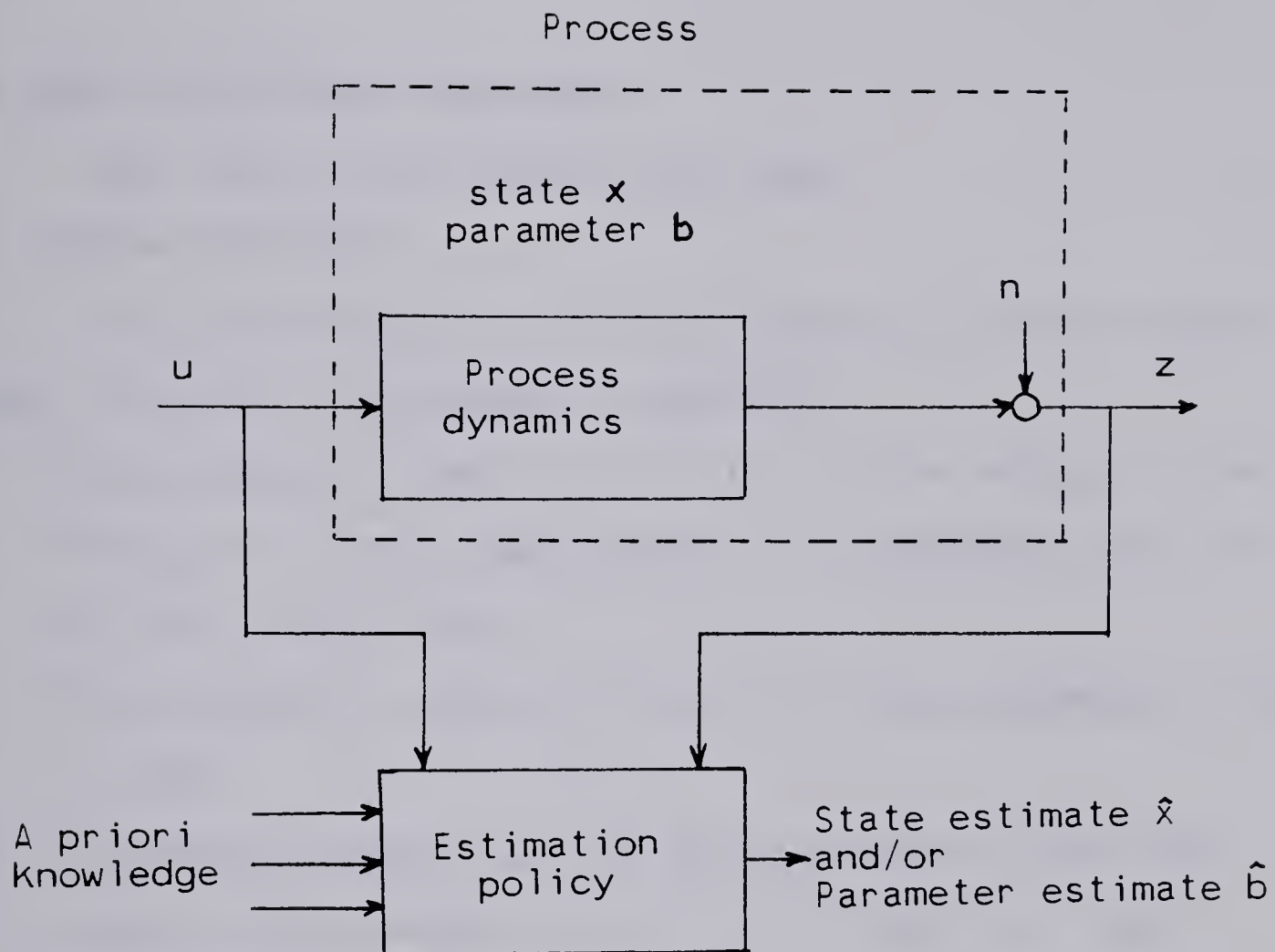


FIGURE IV-1: Engineering formulation of the Identification problem
(from [20])

$$b = b\{u(1), \dots, u(T), z(1), \dots, z(T)\} = b\{\underline{u}', \underline{z}'\} \quad (3)$$

such that a numerical value b (estimate) can be assigned for b .

B. SOME ESTIMATION TECHNIQUES

Some types of estimators are then:

B.1 Bayes Estimator

As a starting point, let us choose a situation where much "a priori" knowledge is available:

1. The probability density function of the noise n ; from this $p(z)$ follows, and since it is dependent on b , it will be noted $p(z/b)$.
2. The probability density function of the parameter values b : $q(b)$
3. The cost of choosing b for the estimate if the true value of the process is b : i.e. a "loss" or "cost" function $C(b, b)$.

Then, from Bayes' rule:

$$p(\underline{z}/\underline{b}) \cdot q(\underline{b}) = p(\underline{z}, \underline{b}) = p(\underline{b}/\underline{z}) \cdot p(\underline{z}) \quad (4)$$

we get

$$p(\underline{b}/\underline{z}) = p(\underline{z}, \underline{b}) / p(\underline{z}) = p(\underline{z}/\underline{b}) \cdot q(\underline{b}) / p(\underline{z}) \quad (5)$$

where

$$p(y) = \int \int \dots \int_{m+1} p(\underline{z}, \underline{b}) \cdot d^{m+1} \underline{b} \quad (6)$$

$p(\underline{b}/\underline{z})$ is the "a posteriori" probability density function of the parameters $\underline{b}$, given the results of the measurements on $\underline{z}$, say $\underline{z} = \underline{c}$.

Then the conditional risk of choosing $\underline{b}(\underline{z})$ if the parameter is $\underline{b}$ can be written as the expectation of the cost function with respect to the observations $\underline{z}$.

The average risk is the expectation with respect to the probability of the value of $\underline{b}$:

$$R(\underline{b}) = E_{\underline{b}} \{E_{\underline{z}/\underline{b}}[C(\underline{b}, \underline{b})]\} \quad (7)$$

The estimate that minimizes this expression is called the "Minimum Risk" or "Bayes" estimate.

We have from Bayes' rule:

$$R(\underline{b}) = \int \int \dots \int d^k \underline{z} \cdot p(\underline{z}) \int \int \dots \int C(\underline{b}, \underline{b}) \cdot p(\underline{b}/\underline{z}) \cdot d^{m+1} \underline{b} \quad (8)$$

$$\min_{\underline{b}} R(\underline{b}) = \min_{\underline{b}} \int \int \dots \int C(\underline{b}, \underline{b}) \cdot p(\underline{b}/\underline{c}) \cdot d^{m+1} \underline{b} \quad (9)$$

since $P(\underline{z}) \geq 0$ for $\underline{z} = \underline{c}$ observed sample.

Then, the necessary condition is:

$$d/d\underline{b} \left(\int \int \dots \int C(\underline{b}, \underline{b}) \cdot p(\underline{b}/\underline{c}) \cdot d^{m+1} \underline{b} \right) = 0 \text{ at } \underline{b} = \underline{b} \quad (10)$$

B.2 Maximum Likelihood Estimator (MLE)

We now drop assumptions 2 and 3. The only "a priori" knowledge left is $p(\underline{z}/\underline{b})$.

Before measurement, we know that the samples of $\underline{z}$ are random variables and have the joint probability function:

$$p\{\underline{z}(1), \dots, \underline{z}(T)/\underline{b}\} \quad (11)$$

After measurement, we know the observed sample values:

$$\underline{z}(1) = \underline{c}, \dots, \underline{z}(T) = \underline{c}_T \quad (12)$$

We write the dependence between $\underline{c}_i$ and $\underline{b}$ as:

$$L \{c_1, \dots, c_r | b\}$$

However, this can also be considered as a function of $\underline{b}$ given the sample and is known as the Likelihood function.

The philosophy of the Maximum Likelihood algorithm is developed in more detail in section IX.

For the sake of convenience, $\text{Log } L$ is usually considered.

Since the logarithm is a monotone function, the maximum of L with respect to b can be obtained by solving:

$$d/d\underline{b} \text{Log} \{L(\underline{c}; \underline{b})\} = 0 \text{ at } \underline{b} = \hat{\underline{b}} \quad (13)$$

Some of the properties of the ML estimator are:

- consistency
- invariance: if $\underline{b}$ is the MLE for $\underline{b}$, $g(\underline{b})$ is the MLE for $g(\underline{b})$.
- asymptotic unbiasedness: $E(\underline{b}) = \underline{b}$ for $t \rightarrow \infty$.
- asymptotic normality: $p(\underline{b}, \underline{b})$ approaches a normal distribution for $t \rightarrow \infty$.
- asymptotic efficiency: approaches the best accuracy or minimum variance as given by the Cramer-Rao inequality.

This inequality states (only the form for an unbiased estimator is given here) that for any estimator $\underline{b}$ of $\underline{b}$:

$$\text{Var}(\hat{\underline{b}}) \geq -1 / E \{ d^2 \text{Log } L / d\underline{b}^2 \} \quad (14)$$

This called the "Minimum Variance Bound". Or, for the multidimensional case:

$$\text{Cov.}(\hat{\underline{b}}) \geq J^{-1} \quad (15)$$

where:

$$J = E \{ d^2 \text{Log } L / d\underline{b} \bullet d\underline{b}' \} \quad (16)$$

J is called the "Fisher Information Matrix" and puts a

bound on the maximum achievable accuracy. A detailed discussion of this topic is provided by Dhrymes [30].

In view of all these properties, the use of ML estimates is highly desirable.

B.3 Markov and Least Squares Estimates

These estimators belong to the class of linear unbiased estimators, i.e. those that satisfy the following relationships:

$$\underline{b} = Q \bullet \underline{z} \quad (17)$$

$$E(\underline{b}) = \underline{b} \quad (18)$$

Given the process:

$$\underline{z} = U \bullet \underline{b} + \underline{n} \quad (19)$$

and the model:

$$\underline{z} = U \bullet \underline{b} \quad (20)$$

where U is the matrix of observed inputs and $\underline{n}$ is the noise, define the error as:

$$\underline{e} = \underline{z} - \underline{z}_m = U \bullet (\underline{b} - \underline{b}) + \underline{n} \quad (21)$$

and the error criterion to minimize, as the quadratic form:

$$E = \underline{e}' R \underline{e} \quad (22)$$

where R is a matrix of weighting coefficients.

Then the minimum is given by:

$$dE / d\underline{b} = 0 \text{ at } \underline{b} = \hat{\underline{b}} \quad (23)$$

$$\underline{b} = (U' R U)^{-1} \bullet U' R \underline{z} \quad (24)$$

This is known as the general "Weighted Least Squares" estimate.

If R is chosen as N^{-1} where $N = E(\underline{nn}')$ is the covariance matrix of the noise process (assumed known),

then:

$$\hat{\underline{b}} = (\underline{U}' \underline{N}^{-1} \underline{U})^{-1} \bullet \underline{U}' \underline{N}^{-1} \underline{z} \quad (25)$$

is known as the Markov or Generalized Least Squares estimate.

If we set $R = I$ (we do not know N), then we get the well known Least Squares estimate:

$$\hat{\underline{b}} = (\underline{U}' \underline{U})^{-1} \bullet \underline{U}' \underline{z} \quad (26)$$

These estimates are unbiased by construction, consistent under certain regularity conditions¹⁹, and have the property of minimum variance: i.e., for any other $\hat{\underline{c}}$

$$\text{Cov}(\hat{\underline{c}}) \geq \text{Cov}(\hat{\underline{b}}) \quad (27)$$

If the noise has a normal distribution, these estimators can be derived from the Maximum Likelihood estimator. In our continuous relaxation of "a priori" knowledge, let us now assume that the only knowledge available is that of $N = E(\underline{nn}')$.

Then:

$$p(\underline{n}) = (1/(2\pi))^{k/2} \bullet (|N|^{-1/2}) \bullet \exp. \{ (-1/2) \bullet (\underline{n}' N^{-1} \underline{n}) \} \quad (28)$$

where $E(\underline{n}) = 0$ and $E(\underline{nn}') = N$.

We can now write:

$$\begin{aligned} \text{Log } p(\underline{z} - \underline{U}\underline{b}) &= -(1/2) \bullet \text{Log} ((2\pi) \bullet |N|) \\ &- (1/2) \bullet [(\underline{z} - \underline{U}\underline{b})' \bullet N^{-1} \bullet (\underline{z} - \underline{U}\underline{b})] \end{aligned} \quad (29)$$

Maximizing this likelihood function leads to:

$$d/d\underline{b} \{ (\underline{z} - \underline{U}\underline{b})' \bullet N^{-1} \bullet (\underline{z} - \underline{U}\underline{b}) \} = 0 \text{ at } \underline{b} = \hat{\underline{b}} \quad (30)$$

or:

$$\hat{\underline{b}} = (\underline{U}' \underline{N}^{-1} \underline{U})^{-1} \bullet \underline{U}' \underline{N}^{-1} \underline{z} \quad (31)$$

¹⁹ See p. 31

which is the Markov estimate.

If in addition we relax knowledge of N , i.e. the "a priori" knowledge is nil, we obtain:

$$\hat{\underline{b}} = (\underline{U}' \underline{U})^{-1} \bullet \underline{U}' \underline{z} \quad (32)$$

the Ordinary Least Squares (OLS) estimator.

Thus, the OLS and the GLS (or Markov) estimators can be derived from the MLE under the assumption of gaussian noise. Under these conditions, all these estimators yield the same estimates and attain Minimum Variance Bound.

The presentation of estimation techniques is limited to this rough outline of the main algorithms. A great number of related techniques have been developed to overcome specific problems that arise in practice in model building (see for example Theil [22], Eykhoff [20], etc...)

C. CURRENT RESEARCH

Contrarily to current research in econometrics, where, because of the wide acceptance and history of practice of the simultaneous equation model, most of the research is directed towards improving on existing estimation techniques or refining the paradigm in the area of nonideal situations, the engineering trends of current research are much more oriented to questioning accepted practices and investigating new theoretical approaches, probably because as they enter the social field with their new tools and their "systems approach", engineers find traditional methods and outlooks inappropriate for the magnitude of the task and the nature of

the problems to solve...

The following few examples are only intended as an illustration of this trend and constitute by no means a review of current research.

C.1 Model Validation - Sensitivity Testing

In their effort to select the appropriate variables to include in their models, most of the econometric model-builders have used single equation statistical testing regardless of whether the system under consideration is recursive or simultaneous, although the limitations of such techniques have been recognized for a long time.

Such tests include the popular "t-test" of parameter significance, the analysis of residuals, the Durbin-Watson statistic and the partial correlation coefficients.

The usefulness of these tests has come to be more and more criticized recently in both the econometric and engineering field.

In particular, Mass and Senge [70] have found that these tests provide information only on the precision with which a given parameter can be estimated from the available data and not on the importance of the associated variable on determining model behavior, that reliance solely on these tests for model specification can lead to the rejection of important parameters on the basis of statistical "insignificance", and they argue that "only tests that examine a variable's influence on overall model behavior

provide a sound basis for assessing an individual variable's importance."

Basing their argument on the use of system dynamics feedback models for social sciences, they then proceed to develop an alternative testing approach of the "model-behavior" kind and provide two examples of application of their technique.

A recent case study by Ford *et al.* [71] supports this approach by a very detailed study of the behavior of large computer models.

This theme is expanded on in the conclusion.

C.2 The use of Information Theory

In a number of fields of physical and engineering science, theorems have been derived with respect to the maximum result obtainable. For example the Carnot cycle in thermodynamics, Heisenberg's relationship in physics , Shannon's maximal channel capacity in communication theory. In the fields of parameter and state estimation, there is a need too for such limiting theorems.

" It would be most useful if it were possible to indicate the maximum amount of information about a parameter or a state vector that could be derived in clearly specified situations. In that way, different estimation techniques or schemes applied to a given situation could be compared just as communication techniques can be evaluated by communication (information) theory." [20]

The comparison of parameter or state estimation and information theory suggests itself very strongly. The values to be estimated are the messages sent by an information source; the transmitter, channel, receiver represent the information processing scheme; the noise stands for the disturbances in process and measurement devices.

Some problem statements have already been made by Zaborsky [72], Caines [73] and others.

The best limiting theorem available at present is the Cramer-Rao bound which provides knowledge of the accuracy that can be obtained under certain prespecified conditions.

A great part of current research, whether dealing with identifiability, estimation algorithms or model structure is moving in the direction of providing such bounds.

Information theoretic methods have been suggested by many authors for the solution of the related problems of hypothesis testing, signal detection and model identification. In recent years, Kullback's information measure [74] and Bhattacharrya's metric have been made popular [75], [76].

The identifiability, model-structure determination and estimation problems are not seen as separate anymore but are being increasingly integrated in one single process. A few examples of this trend are Akaike's "information criterion" (AIC), [77],[78] based on a Maximum Likelihood approach formulated in such a way that it eliminates the need for subjective judgement in hypothesis testing, Rissanen's use

of an entropy measure [79], Tse's quantitative measure of identifiability [80], and Caines' notion of "predictor set completeness" [73]...

C.3 Case Studies

In surveying the actual estimation algorithms used in the determination of the models of interest to the engineering literature, the Maximum Likelihood method strikingly appears as the most used and the most powerful, specially in its "filtering form" which provides the best algorithm to date for the estimation of state-space models with both process and measurement noise. In view of its importance and since it is the main topic of this thesis, the method and a few applications are presented separately in section IX.

V. THE IDENTIFIABILITY PROBLEM

One of the major problems that arise in the estimation of multivariate models is the "identifiability" problem (the "identification" problem in the econometric literature; the control "identification" is equivalent to the econometric "specification").

This can best be illustrated by an example taken from econometrics.

Consider the following simple demand and supply system:

$$\text{Demand: } D_o: q_t = a + b \cdot p_t + u_t \quad (1)$$

$$\text{Supply: } S_o: q_t = c + d \cdot p_t + v_t \quad (2)$$

where q_t is the quantity bought at t , and p_t is the price of the commodity at that time.

Let us take a linear combination of the relations:

$$D_1 = k \cdot D_o + (1-k) \cdot S_o \quad (3)$$

$$S_1 = l \cdot D_o + (1-l) \cdot S_o \quad (4)$$

we get:

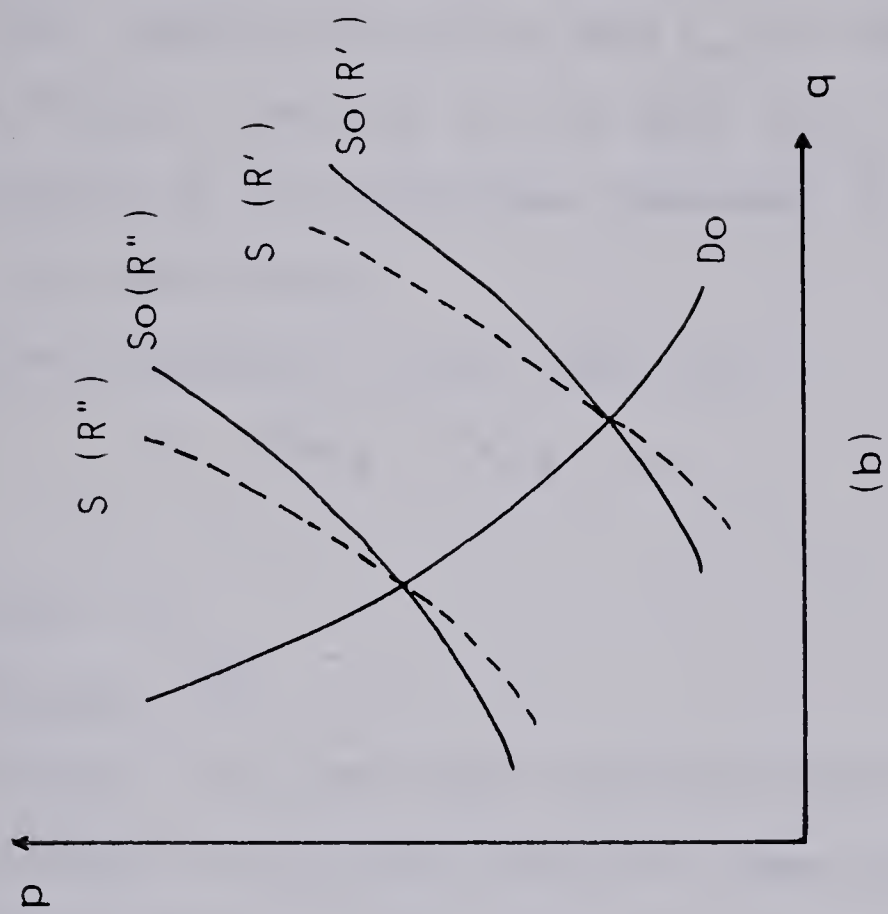
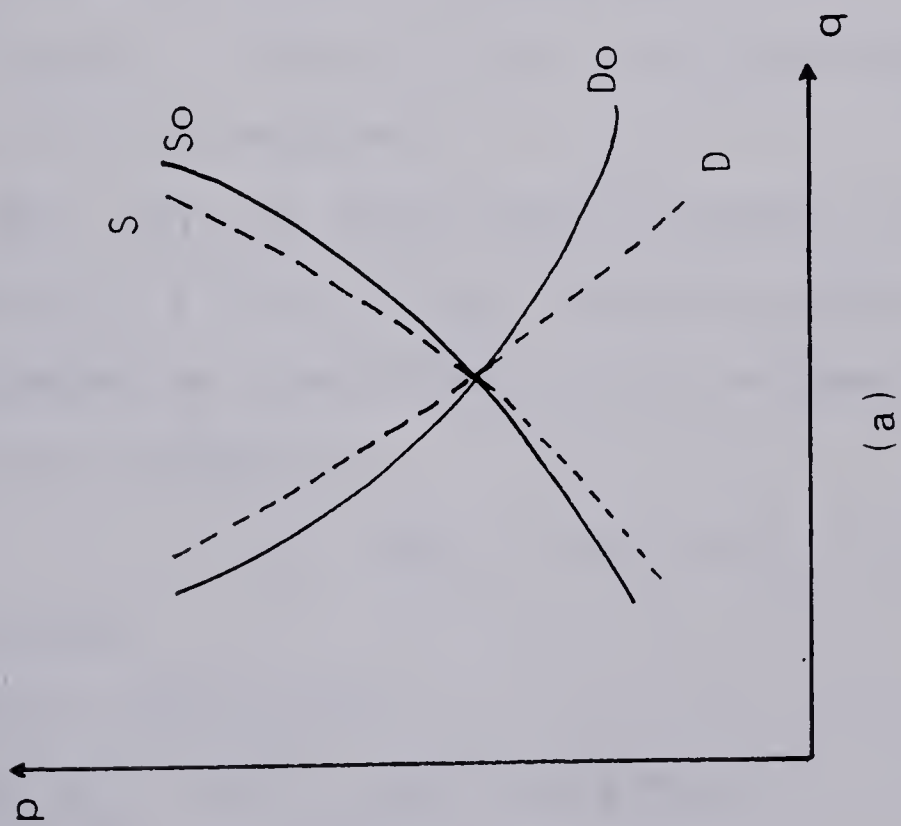
$$\begin{aligned} D_1: q_t &= a \cdot k + c(1-k) + p_t \cdot [(b \cdot k + d \cdot (1-k))] \\ &\quad + k \cdot u_t + (1-k) \cdot v_t \end{aligned} \quad (5)$$

$$\begin{aligned} S_1: q_t &= a \cdot l + c \cdot (1-l) + p_t \cdot [(b \cdot l + d(1-l))] \\ &\quad + l \cdot u_t + (1-l) \cdot v_t \end{aligned} \quad (6)$$

As illustrated in Figure V-1, (D_o, S_o) and (D_1, S_1) are observationally equivalent: that is, given an observation on q_t and p_t , there is no way to decide which is the true model.

Now, if we take for S_o :

$$q_t = c + d \cdot p_t + n \cdot r_t + v_t \quad (7)$$



(a): Both (D_o, S_o) and (D, S) are observationally equivalent.
 (b): If "rainfall" (R' or R'') is specified, Demand (D) is identified and only S_o and S are observationally equivalent

FIGURE V-1: Observationally equivalent models

where r_t is "rainfall" and r_t does not appear in D_o , then, any linear combination of D_o and S_o will not be an equation of the form D_o (because of the term in r_t) and therefore, the equation D_o is identified. However, in this case, S_o is still not identified.

More formally, let our model be:

$$S_1 : B \bullet \underline{y}_t + C \bullet \underline{x}_t = \underline{u}_t \quad (8)$$

with:

$$E(\underline{u}_t) = 0$$

$$E(\underline{u}_t, \underline{u}'_t) = \Sigma$$

And $P(\underline{y}_t / \underline{x}_t)$ the conditional distribution of $\underline{y}_t$ given $\underline{x}_t$. If two different structures yield the same conditional distribution of $\underline{y}_t$, they are called "observationally equivalent" and hence are not identifiable.

To obtain $P(\underline{y}_t / \underline{x}_t)$, we write:

$$P(\underline{y}_t / \underline{x}_t) = P(\underline{u}_t / \underline{x}_t) \bullet |J| = P(\underline{u}_t / \underline{x}_t) \bullet |d\underline{u}_t / d\underline{y}_t| \quad (9)$$

where J is the Jacobian.

Since $d\underline{u}_t / d\underline{y}_t = B$, equation (9) can be written:

$$P(\underline{y}_t / \underline{x}_t) = P\{(\underline{u}_t = B \bullet \underline{y}_t + C \bullet \underline{x}_t) / \underline{x}_t\} \bullet |B| \quad (10)$$

Suppose we premultiply (8) by an nonsingular matrix F to obtain the system S_2 :

$$S_2 : FB \underline{y}_t + FC \underline{x}_t = F \underline{u}_t = \underline{v}_t \quad (11)$$

and we have:

$$E(\underline{v}_t) = F \bullet E(\underline{u}_t) = 0 \quad (12)$$

$$E(\underline{v}_t, \underline{v}'_t) = E(F \underline{u}_t, \underline{u}'_t F') = F \bullet \Sigma \bullet F' \quad (13)$$

Then:

$$P(\underline{y}_t / \underline{x}_t) = P(\underline{v}_t / \underline{x}_t) \bullet |d\underline{v}_t / d\underline{y}_t| \quad (14)$$

where :

$$d\underline{v}_t/d\underline{y}_t = FB \quad (15)$$

and

$$P(\underline{v}_t/\underline{x}_t) = P(\underline{u}_t/\underline{x}_t) \cdot |d\underline{u}_t/d\underline{v}_t| \quad (15)$$

but:

$$d\underline{u}_t/d\underline{v}_t = F^{-1} \quad (17)$$

so that :

$$P(\underline{y}_t/\underline{x}_t) = P(\underline{u}_t/\underline{x}_t) \cdot |F^{-1}| \cdot |FB| = P(\underline{u}_t/\underline{x}_t) \cdot |B| \quad (18)$$

Thus we get the same distribution for $\underline{y}$, and S_1 and S_2 are observationally equivalent.

Econometric theory then goes on defining "admissible" transformations F that yield equivalent systems, and conditions for identifiability.

If the system is written as:

$$A \bullet \underline{z}_t = \underline{u}_t \quad (19)$$

where:

$$A = (B \ C) \quad (20)$$

$$\underline{z}_t = (\underline{y}'_t \ \underline{x}'_t) \quad (21)$$

and if there are R linear homogeneous restrictions on (say) the first row of A , written as:

$$\alpha_i \bullet \phi_i = 0 \quad (22)$$

where:

$$\alpha_i = (a_{i1}, \dots, a_{in}) \quad (23)$$

then, the first equation is "just-identified" if :

$$\text{rank}(A \bullet \phi_i) = G-1 \quad (24)$$

where G is the number of equations.

If $\text{rank}(A \bullet \phi_i) > G-1$ ($< G-1$), the system is over

(under)-identified.

In the latter case, the equation is inestimable. A similar condition exists for the covariance matrix of the disturbances.

Thus, before starting any estimation procedure, one has to check the just(or over) identifiability of each equation, and if needed, add restrictions to make it just(over) identified.

For a more detailed exposition of this problem, see Fisher [81].

Control engineers have independently dealt with this problem (see for example Tse and Anton [82]) using a different terminology and a different approach. A general result for linear systems is that if the minimal dimension of the system is known²⁰ controllability, observability and stability imply identifiability²¹ [82]

Mehra [24] states some additional conditions; in particular, that the system be of minimal order is equivalent to say that only the proper canonical representation is identifiable.

It is for this reason that here, and in a great part of current research, minimal forms of state space equations assume primal importance (see Akaike [77]).

²⁰For the definition of "minimal dimension", a detailed exposition of the relationships between various forms of a state-space model and their relationship to the econometric formulation, see Mehra [24].

²¹ All these concepts derive from modern control theory and are defined in e.g. [83]

It is worth noting that the econometric conditions were formulated in terms of *each* equation whereas in control theory, identifiability of the whole system is examined. The conditions in terms of ranks of matrices amount to the same idea, however.

Tse and Anton [82] also give detailed conditions for the identifiability of parameters in terms of their probability distribution function and the Maximum Likelihood principle.

In practice however, the great majority of the applied literature examines the identifiability of a model only when faced with estimation problems (ill conditioned or singular matrices, physically nonrealizable parameter estimates and large associated error covariances). The usual approach, as outlined by Mehra [84], is then to try fixing some of the parameters, to use a priori weighing for some of the parameters, to try constrained optimization or rank-deficient solutions. All of these methods depend entirely on the modeler's judgement...

VI. EXPERIMENTAL WORK

A. INTRODUCTION

As stated earlier, (see section (I-D)), the purpose of the present work is to investigate the behavior of alternative estimation techniques, drawn from both the econometric and engineering fields, on the Market-Growth model.

Senge's conclusion that the OLS and GLS were inadequate to estimate system dynamics models together with the preceding review of existing econometric and engineering methods show that from the many techniques available, only a few might suit our needs, that is, the estimation of a nonlinear-in-the-variables feedback model in the presence of both equation and measurement noise.

Since the survey of estimation techniques provides no information as to the "a priori" performance of the various estimators in this particular situation, (and even less if we were to decide on the identifiability of these models), it is ultimately experimental work that will give us an idea of their performance. This however will *not* provide a basis for generalization, as pointed out by Senge [17]. It is only when enough knowledge has been accumulated from various sources that ,hopefully, some conclusion can be derived.

Only two of the methods of interest to us were available as estimation packages or subroutines at the University of Alberta, namely the 2SLS (widely used in

econometric work) and the LISREL model. The use of the Filtering form of Maximum Likelihood function, although amply documented by Mehra [84], was not available as a package and therefore a program had to be written in order to use this method²²

The following sections describe the experimental framework and the results and analysis of the various experiments. As a first step, Senge's experiments were repeated using OLS, in order to confirm his results and to have a basis for comparison with the other techniques in the same environment. An alternative econometric method, the LISREL model, has been tried and found unsuccessful. The results are not reported here. A computer program based on Mehra's formulation of the Filtering Form of the Maximum Likelihood algorithm was then developed, tested on lower order, simple models and finally applied to the Market-Growth model.

²² A general PL/I program called GPSIE for "General Purpose System Identifier and Evaluator" was developed by D. Peterson [85] in 1976. It has, among other possibilities, the capacity to compute Full Information Maximum Likelihood (FIML) estimates of parameters in a dynamic state-space model in the presence of equation and measurement noise and is based on the work of Schweppe [86], [87]. This however was not available.

During the course of this research, a later version of this program, closely linked to the DYNAMO, called DYNASTAT was also developed at MIT.

B. THE EXPERIMENTAL FRAMEWORK

The specific details and models for each estimation technique are described in detail in the respective sections. A brief overview of the framework developed by Senge [17] is presented here.

B.1 The experimental procedure

A data generating model (the parameters of which are known) having properties of real-life systems (feedback, nonlinearities, complex dynamics, noise) generates synthetic "perfect" data. This data passes through a measurement process which can reflect imperfections in data measurement by sampling, including a random component or withholding synthetic data for a particular variable to reproduce unavailability of data.

The investigator also specifies a second model with unknown parameters (the estimation model) allowing for specification errors if he wishes to reproduce imperfections in model specification.

The estimation technique under test can then be applied to the estimation model and using synthetic data, yields a set of estimated parameters and statistics for hypothesis testing.

By comparing the estimated parameter values to the values of the "true" system, the investigator can assess the accuracy of the estimation technique. By analysing the statistical test results, he can find out whether or not the tests reliably alert the modeler to inaccuracies. The

process is illustrated in Figure VI-1.

B.2 Variations in experimental conditions

Figure VI-2 illustrates the breadth of experimental conditions which, according to Senge [17], must be investigated separately and in combination in order to test thoroughly the accuracy of the estimation technique under consideration and the effect on estimation results of different aspects of real-life model-building and data measurement.

Figure VI-2 is self explanatory. For a detailed discussion, see Senge [17].

This is not in itself a new idea. This framework is exactly the same as the one developed for Monte-Carlo experiments²³ in the econometric literature except that the method used there are even more sophisticated (as demonstrated for example by the concept of Mean Square Error MSE²⁴). It is quite simplistic to use only "a few" runs to "compare" parameter values and is a weakness in Senge's reporting of results as compared to the econometric literature.

The new idea here is that "realistic" conditions must be met when testing these estimation techniques, which is not the case of most econometric Monte-Carlo experiments as pointed out in section (III-C.5).

In particular, in the present situation, the problem of complex nonlinear, noisy "system dynamics" models that have

²³See section (III-C.5)

²⁴See p. 41

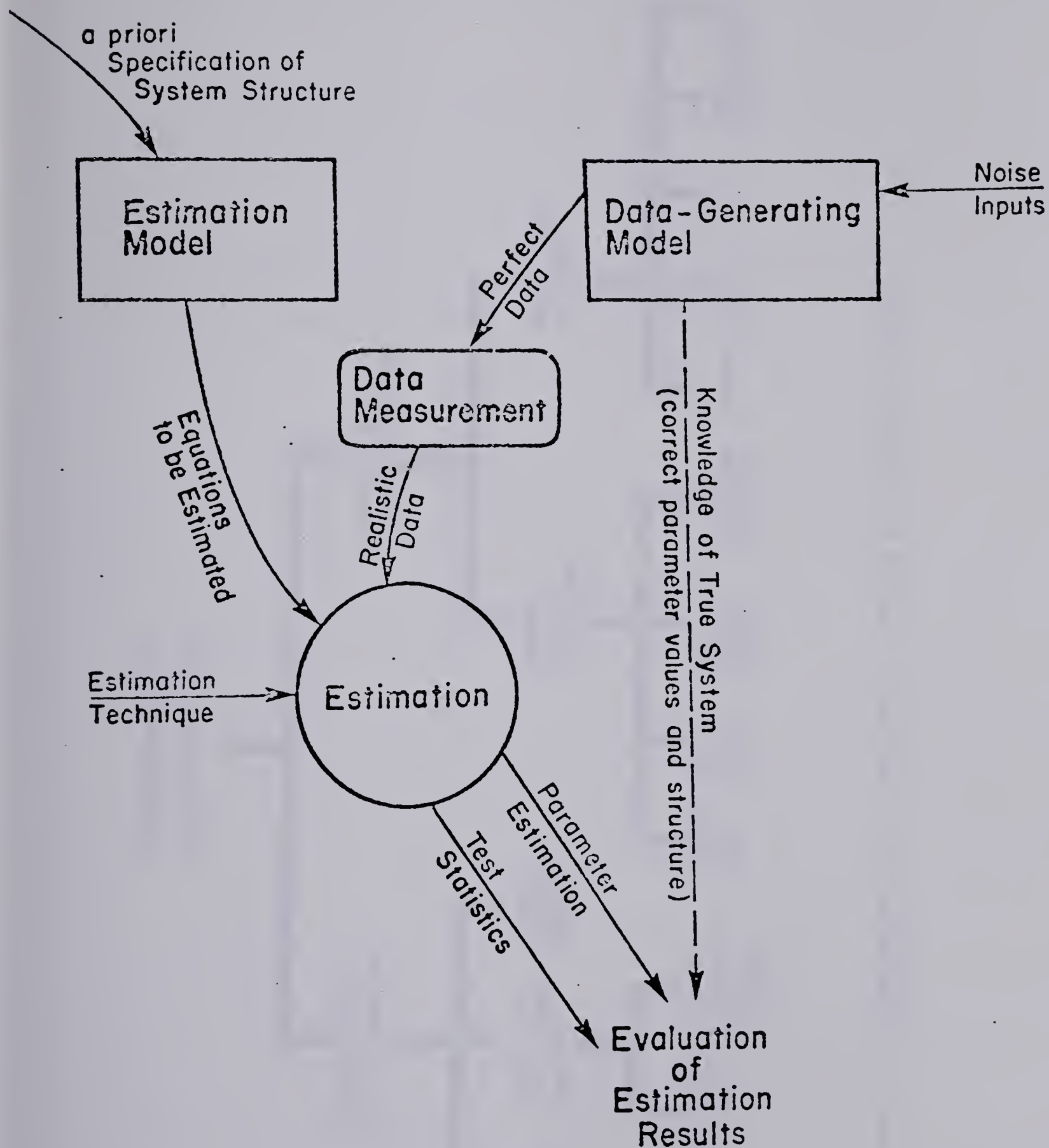


FIGURE VI-1: The experimental procedure
(from [17])

Departures from
Ideal Conditions

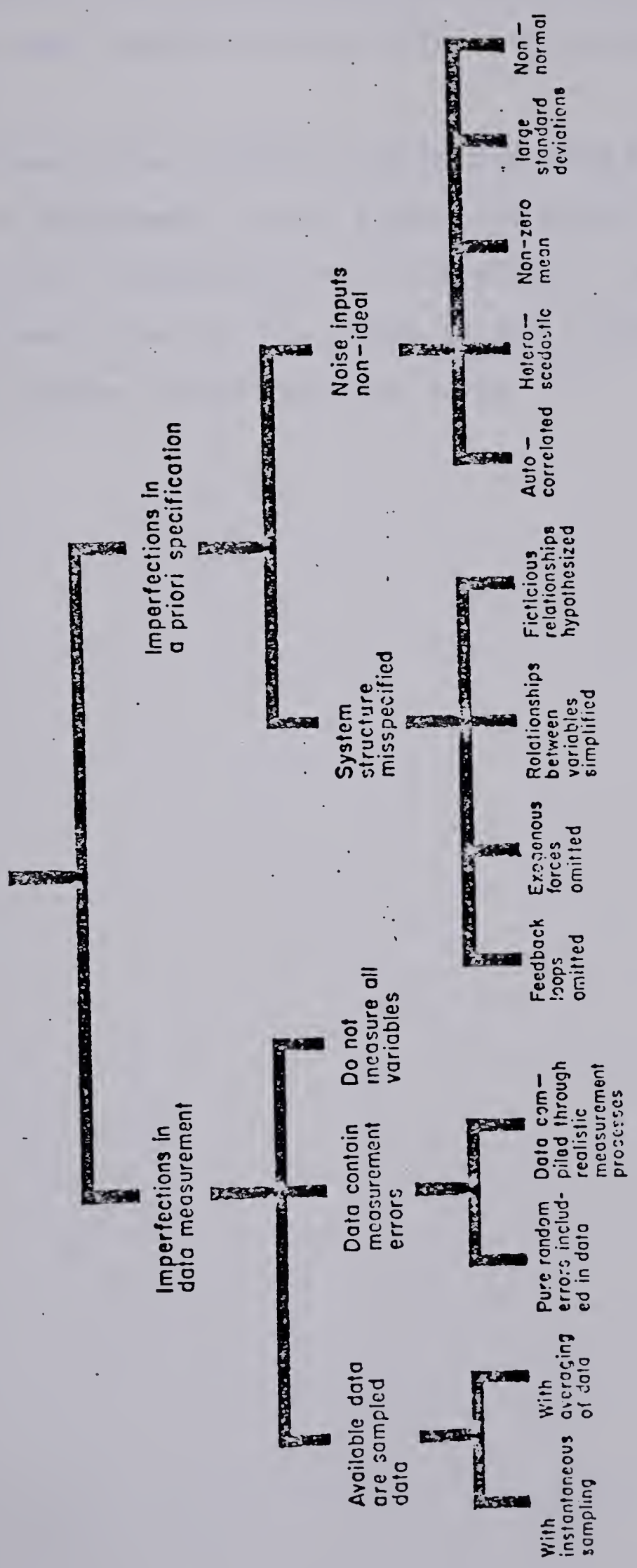


FIGURE VI-2: Possible non-ideal conditions for testing an estimation technique.
(from [17])

not been built with estimation in mind and that therefore reproduce many aspects of real-life, are simultaneously examined.

In view of the difficulties encountered with the problem of measurement noise alone, the present approach was limited to the investigation of its effects, and no attention was given at this stage to the problem of misspecification investigated by Senge [17].

VII. REPRODUCING SENGE'S EXPERIMENTS

A. THE "MARKET GROWTH" MODEL

The "Market-Growth"²⁵ model was developed by Forrester [19] as an illustration of system structure concepts in system dynamics, and shows how a firm's marketing and capital investment policies can produce unstable growth in the market share.

All the relationships necessary to create the growth and stagnation patterns encountered are included within the boundaries of the system. Within the closed boundary²⁶, the system consists of interacting feedback loops as illustrated in Figure VII-1²⁷

Three major feedback loops are shown:

1. LOOP 1 is a positive feedback loop involving the marketing effort: increases in orders booked increase backlog of unfilled orders which increase the delivery rate; since sales are thus boosted, the marketing budget can be increased, which means more salesmen and therefore more orders booked.
2. LOOP 2 involves delivery delay and sales effectiveness: as the backlog increases, delivery delay increases thereby diminishing the firm's marketing effectiveness.
3. LOOP 3 determines the production capacity: rising order backlog, as indicated by delivery delay, is taken as an

²⁵Where growth is the growth of a new product

²⁶ See section (I-A.2)

²⁷ From [17]

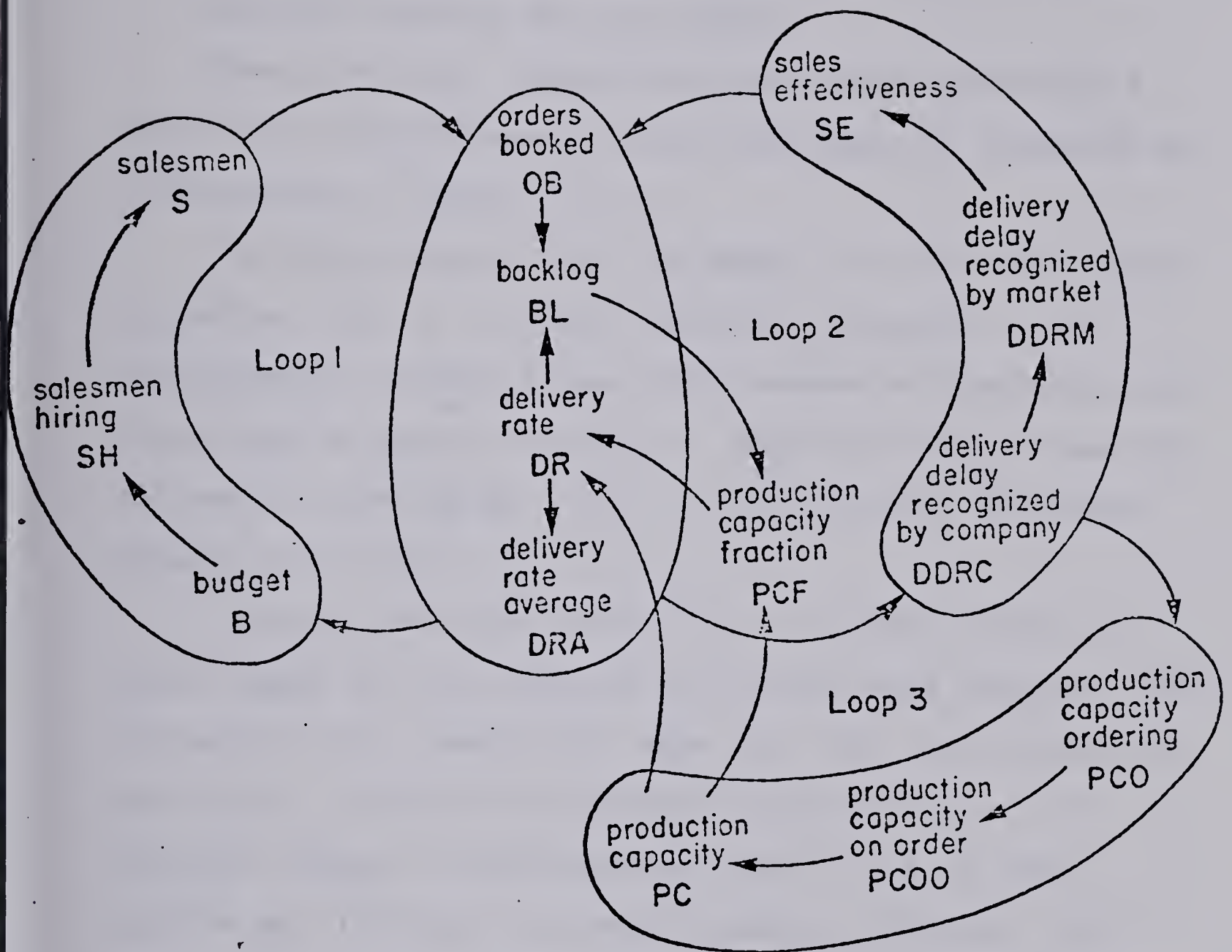


FIGURE VII-1: Major feedback loops in the Market-Growth model.
(from [17])

indication of inadequate production capacity. More capacity is ordered, which, after an acquisition delay, increases production capacity and delivery rate and therefore reduces delivery delay.

When simulated, the Market-Growth model generates a pattern of *unstable* market growth and capacity expansion as illustrated in Figure VII-2.²⁸

The Market-Growth model is modest in size (nine state variables), but it is highly nonlinear (because of the relationships between sales effectiveness and delivery delay recognized by market, production capacity fraction²⁹ and the minimum delivery delay, etc... illustrated by nonlinear tabular functions).

Although the model as written in DYNAMO contains a great number of relationships including Table functions, (see Forrester [19]), Senge [17] shows that for the purposes of estimation, it can be reduced to the following set of discrete linear-in-the parameters form: (All variable symbols are identical to symbols used by Forrester [19], with the exception of the variables PC001, PC002 and PC003 involved in the delay in acquiring production capacity; the three variables are subsumed in the "production capacity on order" PC00 in Forrester)

²⁸ This figure is generated by a DYNAMO run taken from the description in Forrester [19]

²⁹ a multiplier taking into account the variable utilization of production capacity

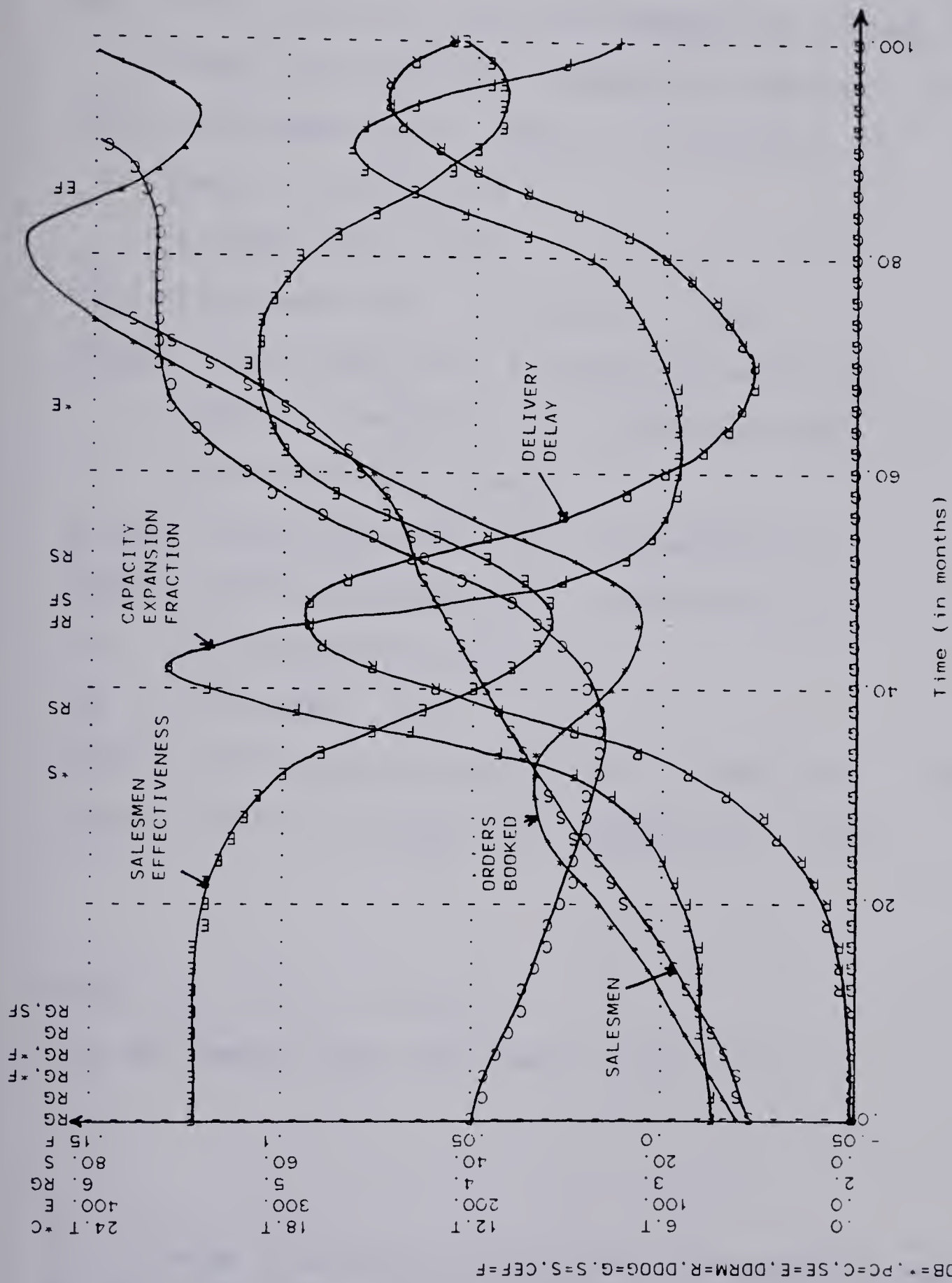


FIGURE VII-2: Behavior of the Market-Growth model

The model equations are:³⁰

$$\begin{aligned} dBL = DT \bullet [& K1 \bullet S(-DT) + K2 \bullet S(-DT) \bullet DDRM(-DT) + K3 \bullet BL(-DT) \\ & + K4 \bullet BL(-DT)^2 / PC(-DT) + K5 \bullet BL(-DT)^3 / PC(-DT)^2 + N1] \end{aligned} \quad (1)$$

$$\begin{aligned} dDRA = DT \bullet [& K8 \bullet BL(-DT) + K9 \bullet BL(-DT)^2 / PC(-DT) + \\ & K10 \bullet BL(-DT)^3 / PC(-DT)^2 \\ & + K11 \bullet DRA(-DT) + N2] \end{aligned} \quad (2)$$

$$dS = DT \bullet [K12 \bullet DRA(-DT) + K13 \bullet S(-DT) + N3] \quad (3)$$

$$\begin{aligned} dPC001 = DT \bullet [& K14 \bullet PC(-DT) + K15 \bullet PC(-DT) \bullet DDC(-DT) \\ & + K16 \bullet PC(-DT) \bullet DDC(-DT)^2 + K17 \bullet PC(-DT) \bullet DDC(-DT)^3 \\ & - 0.25 \bullet PC001(-DT) + N4] \end{aligned} \quad (4)$$

$$dPC002 = DT \bullet [0.25 \bullet PC001(-DT) - 0.25 \bullet PC002(-DT)] \quad (5)$$

$$dPC003 = DT \bullet [0.25 \bullet PC002(-DT) - 0.25 \bullet PC003(-DT)] \quad (6)$$

$$dPC = DT \bullet [0.25 \bullet PC003(-DT)] \quad (7)$$

$$DDC = (1/2) \bullet DDRC - 0.3 \quad (8)$$

$$dDDRC = DT \bullet [0.25 \bullet [BL(-DT) / DRA(-DT) - DDRC(-DT)] + N9] \quad (9)$$

$$dDDRM = DT \bullet [(1/6) \bullet [DDRC(-DT) - DDRM(-DT)] + N10] \quad (10)$$

Where: $dBL = BL - BL(-DT), \dots$

and the symbols are explained in Table VII-1.

³⁰ In Senge's documents [17], [18] this equation appears without the term " $-0.25 \bullet PC003(-DT)$ ". This is assumed to be an error since the proper transcription of a third-order delay is as indicated here (see [12]) and since without this term, the model explodes instead of generating the expected behavior)

TABLE VII-1
MARKET-GROWTH MODEL SYMBOLS

$BL(t)$	order backlog (units)
$DRA(t)$	delivery rate average (units/month)
$S(t)$	number of salesmen (men)
$PC(t)$	production capacity (units/month)
$PC001(t), PC002(t), PC003(t)$	production capacity on order (units/month) ^{3 1}
$DDRC(t)$	delivery delay recognized by company (months)
$DDRM(t)$	delivery delay recognized by market (months)
$DDC(t)$	delivery delay condition

^{3 1}third order delay: see
Forrester [12])

When the Market-Growth model is simulated as the data generating model "DT" in equations (1) to (10) is the simulation interval for the experiment, and N_i denotes the noise input in each equation. When the model serves as the estimation model, DT is the sampling interval of available data, and N_i denotes the implicit error process in each equation.

The parameters to be estimated in the estimation experiments are designated K_i in equation (1) to (4). The exact values of the parameters (that appear in Tables VII-2 to VII-11) are taken from Senge [17]. They often appear with complex coterms (as in equations (1) or (2)) due to the series expansion (up to the third order terms) of Forrester's nonlinear tabular functions. Senge chose the parameterizations which in his estimation best drew statistically significant information from the data. (see [17] footnote 11).

B. EXPERIMENTAL RESULTS

B.1 Ideal conditions

The purpose of the following runs is to demonstrate that in principle OLS can yield accurate results for the Market-Growth model (the specification of which is very different from that of traditional econometric models). That is, that despite the many nonlinearities, feed-back loops, and the formulation of the model, the parameters of the Market-Growth model can be estimated by traditional

econometric means provided that the specification of the model is right and that the data is free from measurement errors.

1. Generation of the data

For these ideal conditions, small noise inputs are added to equations (1),(2),(3),(4),(9) and (10).

Equation (8) is merely an identity and (5),(6) and (7) are restatements of the third order delay in Forrester's model for the purposes of estimation: therefore they do not have noise inputs.

The noise sequences are generated as:

$$N = \text{GGNOF}(\text{*****}) * 0.01 * \text{MX}$$

where GGNOF is a random noise generator from the IMSL library³² [88], ***** is a seed number determining the actual sequence of random numbers and MX is the mean of the left hand side variable computed from noiseless data. GGNOF then generates a normally distributed noise sequence with standard deviation equal to $0.01 * \text{MX}$.

Seven runs were tried with different noise sequences. The seed number is the same for all the inputs in run 1 to 5. Each equation has a different seed number in runs 6 and 7. In both cases the correlation matrix of the six noise sequences for each run indicates that the hypothesis of no cross-correlation between noise sequences is very acceptable (maximum correlation of .2)

³²GGNOF has been deleted from version 7 Of the IMSL library

The simulation interval DT was taken as 1 (month) in all runs. The initial conditions for each level variable are from Forrester [19]; PC001, PC002, PC003 are assumed to have the same initial values as PC00 in Forrester's model. One hundred data points were generated for each run. All level variables are measured (although only BL, DRA, S, PC001, PC, DDC, DDRM are needed for estimation). The experimental conditions for each run are summarised in Table VII-2.

2. Estimation of parameters

The Time Series Processor (TSP) package [89] available at the University of Alberta was used for all OLS estimations.

All variables are lagged once by the estimation program thus leaving 99 points of data for the actual estimation.

The four equations with unknown parameters (equations (1) to (4)) are estimated both in the level variable form and in the first difference form as follows:

$$X(t) = X(t - DT) + \sum_i b_i \bullet X_i(t - DT) \quad (11)$$

$$X(t) - X(t - DT) = \sum_i b_i \bullet X_i(t - DT) \quad (12)$$

In both cases all the parameter values are the same. The only difference is the value of the measure of goodness of fit R^2 (coefficient of goodness of fit) which is much lower when the equations are estimated in the first difference form, as explained by Maddala [33, p. 92]. In the case of equation (8) where the coefficient of

TABLE VII-2
EXPERIMENTAL CONDITIONS FOR TABLES VII-3 to VII-6
IDEAL SITUATION - 1% Equation noise

* Initial conditions for state variables in Market Model

BL = 8000.	DDC = .7	PC001 = -1728
S = 10.	DDRC = .7	PC002 = -1728.
DDRM = 2.	DRA = 4000.	PC003 = -1728.
PC = 12000.		

DT = 1 (month)

* Means of variables used in computing noise levels

MBL	= 40558.5	MDDRC	= 3.29
MDRA	= 13638.5	MDDRM	= 3.24
MS	= 56.529	MPC	= 15589.7
MPC001	= 1499.26		

* Noise Sequences: seed numbers for GGNDF routine

Run #	Seed numbers	for
1	12345	all
2	54321	all
3	29875	all
4	947321	all
5	33674	BL
	2556	DRA
	1212	S
	95642	PC001
	4680	DDRC
	48801	DDRM
6	947312	BL
	2345	DRA
	6312	S
	95642	PC001
	44872	DDRC
	443521	DDRM

100 data points are generated. 99 are available for estimation
There is no measurement noise.

PC001(-1) is not estimated (it is equal to .25 because it is introduced by the rewriting of the third-order delay), the regressions are:

$$PC001(t) - 0.25 \cdot PC001(t - DT) = \sum_i b_i \cdot X_i(t - DT) \quad (13)$$

$$PC001(t) - 1.25 \cdot PC001(t - DT) = \sum_i b_i \cdot X_i(t - DT) \quad (14)$$

For each equation, the regression coefficients are given by:

$$\underline{b} = (Z'Z)^{-1} \cdot Z' \underline{y} \quad (15)$$

where

$\underline{b}$ = vector of regression coefficients (Nx1)

Z = matrix of right-hand side variables (TxN)

$\underline{y}$ = left-hand side variable (Tx1)

Among other output, the TSP package OLSQ routine gives:

- a. The t-statistics for the null hypothesis that individual regression coefficients are zero:

$$t_i = b_i / s_i \quad (16)$$

where s is the estimated standard error for each coefficient.

A coefficient is accepted as being different from zero at the 95% confidence level if the t-statistic is greater than 2. Otherwise, it is "insignificant" and the usual econometric practice is to run another regression without the explanatory variable associated with that coefficient.

- b. The coefficient of goodness of fit R^2 :

$$R^2 = 1 - (\underline{e}' \cdot \underline{e} / (\underline{y} - \bar{y})' \cdot (\underline{y} - \bar{y})) \quad (17)$$

where the residuals $\underline{e}$ are calculated from:

$$\underline{e} = \underline{y} - Z\underline{b} \quad (18)$$

and $\bar{y}$ is the sample mean of y .

The goodness of fit R^2 measures the amount of variation of the dependent variable "explained" by the set of regressors. The closer it is to one, the better the fit.

c. The Durbin-Watson statistic DW:

$$DW = \sum_{t=2}^T \{e(t) - e(t-1)\}^2 / \sum_{t=1}^T e^2(t) \quad (19)$$

where T is the number of data points.

The Durbin-Watson statistic is a measure of the amount of autocorrelation in the least squares residuals (provided that the residuals are normally distributed). If this autocorrelation is equal to 1, $DW = 0$; if it is 0, $DW = 2$, and if it is -1, $DW = 4$.

Hence, if DW is close to 0 or 4, we know the residuals $e(t)$ and $e(t-1)$ are highly correlated. A value close to 2 is indicative of little or no autocorrelation in the residuals (noise sequence).

3. Estimation results

The results of seven runs for the ideal conditions are reported in Tables VII-3 to VII-6. t -statistics appear in parenthesis below each estimated coefficient and the values of R^2 , in level and first difference

TABLE VII-3

Market model - Ideal conditions

OLS estimates of parameters in equation 1.

Run Number	K1	K2	K3	K4	K5	R ² (R ²)	DW
True Value	475.	-61.5	-0.6178	0.1324	-0.00975	1.0000 (0.9994)	0.15
Senge's Results	486.8 (222.8)	-61.1 (-327.7)	-0.6686 (-164.5)	0.1440 (159.4)	-0.01066 (-126.5)	1.0000 0.9994	0.15
1	348.3 (9.4)	-51.6 (-18.4)	-0.4467 (-5.81)	0.0919 (5.23)	-0.00639 (-4.3)	0.9990 (0.8328)	1.20
2	408.8 (10.8)	-62.1 (-21.3)	-0.5250 (-7.3)	0.1005 (6.45)	-0.00591 (-4.3)	0.9992 (0.8708)	1.20
3	355.2 (8.66)	-46.4 (-16.7)	-0.4587 (-6.0)	0.1023 (5.7)	-0.00803 (-4.44)	0.9991 (0.7719)	1.17
4	325.7 (8.25)	-48.7 (-16.6)	-0.4104 (-5.35)	0.0848 (5.04)	-0.00587 (-4.07)	0.9991 (0.8015)	1.07
5	348.4 (9.86)	-49.9 (-20.4)	-0.4743 (-6.0)	0.1052 (5.2)	-0.00796 (-4.1)	0.9992 (0.8446)	1.29
6	486.5 (13.28)	-62.6 (-24.3)	-0.6986 (-9.2)	0.1357 (7.58)	-0.00875 (-5.6)	0.9991 (0.8802)	1.44
7	311.7 (7.82)	-49.6 (-16.8)	-0.3523 (-4.1) [*]	0.0622 (2.9) [*]	-0.00321 (-1.6) [*]	0.9989 (0.8120)	1.22
Average	369.2	-53.0	-0.4809	0.0975	-0.00659		

TABLE VII-4

Market model - Ideal conditions

OLS estimates of parameters in equation 2.

Run Number	K6	K7	K8	K9	R ² (R ²)	DW
True value	0.6178	-0.1324	0.00975	-1.000		
Senge's Results	0.6317 (15.44)	-0.1335 (-15.72)	0.00962 (15.58)	-1.029 14.87	0.9996 (0.5070)	1.97
1	0.6235 (22.97)	-0.1333 (-23.17)	0.00979 (22.98)	-1.011 (-22.4)	0.9998 (0.8835)	2.08
2	0.5553 (16.54)	-0.1190 (16.62)	0.00875 (16.36)	-0.897 (-16.03)	0.9995 (0.7761)	2.06
3	0.5978 (15.24)	-0.12274 (-15.23)	0.00930 (14.57)	-0.971 (-14.77)	0.9994 (0.7567)	1.87
4	0.6482 (18.27)	-0.1390 (-18.4)	0.01026 (18.4)	-1.050 (-17.8)	0.9997 (0.8050)	2.00
5	0.6539 (19.4)	-0.1393 (-19.4)	0.01018 (18.9)	-1.063 (-18.9)	0.9997 (0.8242)	2.16
6	0.6251 (22.27)	-0.1345 (-22.2)	0.00993 (21.75)	-1.009 (21.75)	0.9997 (0.8570)	2.23
7	0.6093 (18.7)	-0.1301 (-18.8)	0.00953 (18.4)	-0.989 (-18.2)	0.9997 (0.8228)	2.12
Average	0.6163	-0.1318	0.00968	-0.999		

TABLE VII-5

Market model - Ideal conditions

OLS estimates of parameters in equation 3.

Run Number	K10	K11	R ² (R ²)	DW
True Value	3.000x10 ⁻⁴	-0.05		
Senge's Res ults	3.288x10 ⁻⁴ (6.58)	-0.0575 (-4.86)	0.9996 (0.5070)	1.94
1	3.427x10 ⁻⁴ (7.27)	-0.0596 (-5.42)	0.9997 (0.5478)	1.95
2	2.701x10 ⁻⁴ (6.11)	-0.0422 (-4.08)	0.9996 (0.5260)	1.80
3	2.818x10 ⁻⁴ (4.56)	-0.0466 (-3.18)	0.9994 (0.4002)	2.09
4	2.313x10 ⁻⁴ (5.52)	-0.0342 (-3.46)	0.9997 (0.5014)	1.85
5	3.083x10 ⁻⁴ (6.00)	-0.0532 (-4.36)	0.9997 (0.5178)	2.13
6	4.410x10 ⁻⁴ (8.70)	-0.0816 (-7.00)	0.9997 (0.5415)	2.20
7	3.365x10 ⁻⁴ (6.72)	-0.0639 (-5.08)	0.9996 (0.5098)	2.07
Average	3.16x10 ⁻⁴	-0.0545		

TABLE VII-6
Market model - Ideal conditions
OLS estimates of parameters in equation 4.

Run Number	K12	K13	K14	K15	R ² (R ²)	DW
True Value	-0.0698	0.1244	-0.08138	0.02704		
Senge's Results	-0.0752 (-35.39)	0.1366 (28.09)	-0.08983 (-25.71)	0.02837 (36.25)	1.0000 (0.9960)	2.12
1	-0.0703 (-30.52)	0.1257 (25.79)	-0.08243 (25.17)	0.02730 (38.86)	1.0000 (0.9996)	1.92
2	-0.0706 (-23.03)	0.1272 (18.71)	-0.08445 (-18.1)	0.02791 (27.74)	1.0000 (0.9994)	2.58
3	-0.0684 (-21.2)	0.1213 (16.85)	-0.07923 (-15.67)	0.02660 (23.47)	0.9999 (0.9990)	2.11
4	-0.0666 (-25.6)	0.1172 (20.6)	-0.07648 ^a (-19.7)	0.02600 (30.93)	1.0000 (0.9995)	2.06
5	-0.0704 (-26.6)	0.1256 (22.33)	-0.08243 (-21.7)	0.02733 (33.5)	1.0000 (0.9995)	2.16
6	-0.0693 (-28.52)	0.1241 (23.65)	-0.08194 (-23.00)	0.02732 (35.83)	1.0000 (0.9997)	2.13
7	-0.0681 (-28.2)	0.1212 (23.35)	-0.07974 (-22.5)	0.02682 (35.05)	1.0000 (0.9996)	2.12
Average	-0.061	0.1232	-0.08096	0.02704		

forms, and of DW are given for each equation.

As shown in these tables, in principle, the model can be estimated by OLS. This is the only case where all the assumptions of OLS are met:

- 1) $E(u_i) = 0$ for all i : zero mean for the noise
- 2) $V(u_i) = S^2$ for all i : common variance
(homoscedastic process)
- 3) $E(u_i, u_j) = 0$ for $i \neq j$: uncorrelated noise
- 4) $E(u_i, x_j) = 0$ for all i, j : noise uncorrelated
with X

Parameter estimates deviate only slightly from the true parameter values, estimates are uniformly significant at the 99 % confidence level (all t-statistics are above 2.37), R^2 is close to one in all equations and DW close to 2 indicates little autocorrelation in all equations except equation (1). In this case, DW close to one indicates about 50% autocorrelation which according to Senge [17] stems from the linearization of the nonlinearities in that equation. However, the values of K_1 to K_5 are consistently worse than those indicated by Senge [17], possibly due to this autocorrelation. All other estimates are comparable to or better than those given by Senge.

4. Conclusion.

Although only a few runs were attempted in order to reproduce Senge's "typical" results, under these ideal conditions, even though the model is highly nonlinear

and includes many feedback loops, thus differing notably for usual econometric formulation, the results indicate that OLS is in principle capable of yielding acceptable parameter estimates of the Market-Growth model.

B.2 Imperfections in data Measurement

Senge [17] argues that realistic imperfections in data measurement (10% error is characteristic of most econometric time series; see Morgenstern [90]) result in highly inaccurate estimates of the parameters of the Market-Growth model taken as a specimen of "System Dynamics" feedback models. The following few runs attempt to reproduce some of his results.

1. Experimental conditions.

Uncorrelated random noise, with variance equal to 10% of the mean of each measured level variable , BL, DRA, S, PC001, PS, DDRC, DDRM, (DDC = 1/2*DDRC - 0.3), is now added to these variables.

All the other experimental conditions remain identical to those presented in section (VII-B). Although all the assumptions of OLS are not met in this experiment, OLS estimation is still used here.

The noise inputs are added to the relevant variables as:

$$X'_i = X_i + V_i \quad (20)$$

where:

$$V_i = \text{GGNOF}(\text{*****}) * 0.1 * \text{MX}_i \quad (21)$$

Table VII-7 summarises the experimental conditions for the six runs.

TABLE VII-7
EXPERIMENTAL CONDITIONS FOR TABLES VII-8 to VII-11
10% MEASUREMENT ERRORS

- * Initial conditions same as in Table VII-2
- * Means of variables same as in Table VII-2
- * Noise Sequences: seed numbers for GGNDF routine

Run #	Equation noise	Measurement noise	for
1	as in previous run 5	66751 93023 10735 76925 5675 22768 31356	BL DRA S PC001 DDRC DDRM PC
2	as in previous run 6	66751 42021 10735 16305 5653 1763 39804	BL DRA S PC001 DDRC DDRM PC
3	as in run 1	as in run 2	all
4	as in run 2	as in run 1	all
5	947321	947321	all

2. Estimation results.

The results in Tables VII-8 to VII-11 indicate the severe deterioration of the estimates due to the modest 10% measurement noise.

Statistically significant estimates in equations (1) to (3) deviate from their true value by factors ranging from 2 to 9.

- Equation (1) is the most accurately estimated: all the parameters in this equation fall within a factor of two of the real values and most are statistically significant at the 99% confidence level (t-statistics exceed 2.37)

- Except for the odd discrepancy, most parameters in equation (2) are also within a factor of two to three of their real values, are significant at the 99% probability level and are on the average much better than those indicated by Senge for that equation.

- The most inaccurate results occur in equation (3) where the statistically significant estimates of K_{10} are in error by a factor of 4 (in average) and those of K_{11} , by a factor of 5.

- All estimates in equation (4) have the wrong sign and are statistically insignificant at the 95% level. Thus the modeler is warned of the error; however, he has no way of improving on the specification since it is already a perfectly specified model.

TABLE VII-8
Market model - 10% measurement noise.
OLS estimates of parameters in equation 1.

Run Number	K1	K2	K3	K4	K5	R ² (R ²)	DW
True Value	475.	-61.5	-0.6178	0.1324	-0.00975		
Senge's Res ult	252.8 (3.64)	-31.6 (-3.4)	-0.2942 (-2.24)	0.0660 (1.86)	-0.00586 (-1.54)	0.989 (0.184)	2.12
1	221.2 (2.48)	-26.8 (-1.82)	-0.4151 (-1.92)	0.1123 (1.72)	-0.0100 (-1.43)	0.9582 (0.0693)	2.57
2	238.2 (3.19)	-39.8 (-2.71)	-0.3242 (-2.18)	0.0675 (2.50)	-0.00308 (-2.15)	0.9555 (0.1116)	2.63
3	260.4 (2.89)	-35.0 (-2.24)	-0.4882 (2.30)	0.1337 (2.3)	-0.0110 (-2.02)	0.9591 (0.0903)	2.62
4	294.8 (3.31)	-38.8 (-2.64)	-0.5718 (-2.61)	0.1551 (2.40)	-0.01273 (-1.95)	0.9560 (0.1159)	2.62
5	216.9 (3.06)	-20.1 (-1.58)	-0.30929 (-2.56)	0.0447 (2.20)	-0.0057 (-2.30)	0.9735 (0.0949)	2.45
Average	246.2	-32.1	-0.4204	0.1027	-0.00768		

TABLE VII-9
Market model - 10% measurement noise.
OLS estimates of parameters in equation 2.

Run Number	K6	K7	K8	K9	R ² (R ²)	DW
True value	0.6178	-0.1324	0.00975	-1.000		
Senge's Res ults	0.0876 (2.29)	-0.0246 (-2.69)	0.00227 (2.60)	-0.1024 (-1.54)	0.989 (0.073)	2.13
1	0.4513 (6.69)	-0.1017 (-5.14)	0.00782 (3.68)	-0.7032 (-6.96)	0.9357 (0.3421)	2.47
2	0.1853 (4.39)	-0.0280 (-3.8)	-0.00119 (2.97)	-0.3427 (-4.38)	0.9319 (0.1826)	2.68
3	0.3088 (5.49)	-0.0679 (-4.5)	0.00524 (3.58)	-0.4885 (-5.62)	0.9453 (0.2607)	2.62
4	0.4382 (6.29)	-0.0962 (-4.83)	0.00715 (3.52)	-0.6913 (-6.64)	0.9233 (0.3228)	2.50
5	0.2225 (4.63)	-0.0324 (-4.26)	0.00103 (4.06)	-0.4206 (-4.60)	0.9596 (0.1881)	2.65
Average	0.3212	-0.0652	0.00449	-0.5293		

TABLE VII-10
Market model - 10% measurement noise.
OLS estimates of parameters in equation 3.

Run Number	K10	K11	R ² (R ²)	DW
True Value	3.000x10 ⁻⁴	-0.05		
Senge's Res ults	5.11x10 ⁻⁴ (3.06)	-0.098 (-2.53)	0.993 (0.089)	2.04
1	11.74x10 ⁻⁴ (4.19)	-0.255 (-3.93)	0.9652 (0.1403)	2.74
2	11.11x10 ⁻⁴ (3.77)	-0.241 (-3.56)	0.9586 (0.1132)	2.76
3	10.66x10 ⁻⁴ (3.68)	-0.231 (-3.43)	0.9638 (0.1072)	2.77
4	11.97x10 ⁻⁴ (4.22)	-0.260 (-3.99)	0.9601 (0.1422)	2.73
5	12.73x10 ⁻⁴ (3.71)	-0.289 (-3.51)	0.9444 (0.1167)	2.75
Average	11.64x10 ⁻⁴	-0.255		

TABLE VII-11

Market model - 10% measurement noise.

OLS estimates of parameters in equation 4.

Run Number	K12	K13	K14	K15	R ² (R ²)	DW
True Value	-0.0698	0.1244	-0.08138	0.02704		
Senge's Results	-0.0345 (2.10)	0.0364 (0.93)	-0.01330 (-0.46)	0.01049 (1.56)	0.995 (0.682)	1.85
1	0.0216 (0.94)	-0.0941 (-1.74)	0.0796 (1.98)	-0.00929 (-1.00)	0.9916 (0.8834)	2.50
2	0.0009 (0.03)	-0.0402 (-0.65)	0.0403 (0.92)	-0.00150 (-0.16)	0.9919 (0.8810)	2.05
3	0.0029 (0.10)	-0.0429 (-0.67)	0.0424 (0.92)	-0.00244 (-0.24)	0.9913 (0.8657)	2.04
4	0.0292 (1.24)	-0.1147 (-2.13)	0.09517 (2.46)	-0.01247 (-1.45)	0.9924 (0.8989)	2.55
5	0.0136 (0.47)	-0.0675 (-1.09)	0.05870 (1.41)	-0.0055 (-0.62)	0.9898 (0.8449)	2.28
Average	0.0136	-0.0719	0.0632	-0.06240		

3. Conclusion.

Although Senge reports results of GLS estimation, which he finds much better than the OLS ones, the OLS results obtained here for equations (1) and (2) are on the average much better than Senge's thus putting in doubt the necessity to use GLS. The results for equation (3) are worse than his, however. Even though all the coefficients in equation (4) have the wrong sign here, their statistical significance is comparable to that obtained by Senge: in both cases the modeler is informed that the values are wrong or insignificant. Mass and Senge [70] discuss in detail the fallacy of viewing measures such as the t-test as tests of specification.³³ These results, although inextensive, prove beyond doubt Senge's claim that it is unrealistic and meaningless to try to estimate system dynamics model using such methods as OLS given that all available data contains at least 10% measurement error.

Senge goes on investigating the effect of the relative size of the measurement and equation noise inputs. He finds that increasing the equation noise to 10%, as might be suggested by econometric theory³⁴, does not help improve the accuracy of the estimates (in the contrary some are deteriorated).

Considering that the experimental conditions are still

³³ See section (IV-C.1)

³⁴ See [17]

very ideal (all data available, 100 data points and perfectly specified model), he further demonstrates that a realistic specification error in equation (1) results in "meaningless" estimates... However, since we have limited ourselves here to the investigation of the effect of measurement noise on the estimates and their possible improvement by alternative methods, we will be considering only measurement errors in the sequel.

VIII. ALTERNATIVE ECONOMETRIC METHODS

A. INTRODUCTION

Of all the econometric methods, the most used is the Two-Stage Least Squares technique (2SLS) which was developed specifically to avoid the shortcomings of OLS in "simultaneous" equations systems. Its behavior has been thoroughly investigated and it performs well in most econometric problems, even in the case of dynamic models with lagged variables on the right-hand side of the equations.

This single equation method would seem to have been able to perform at least better than OLS with a dynamic model. It has been used by Mehra [24] on a state-space model to determine model structure prior to the estimation of parameters using Maximum Likelihood.

Unfortunately, the nature of System dynamics models, where all right-hand side variables are lagged, and where, all the behavior of the model being generated within the boundaries, there are no external inputs, rules out the use of 2SLS.

Referring back to section (III-B.2), the 2SLS method goes around the problem of OLS bias by first estimating the reduced form³⁵ of the system by OLS (which yields an unbiased estimate), and then, by replacing the endogenous variables in the right-hand side of equation *i* by their

³⁵ see p. 32

estimates obtained from the reduced form. Since system dynamics models are already in the reduced form, (there are no contemporaneous endogenous variables in the right-hand side of the equations), the 2SLS method is not applicable.

This situation illustrates very well the debate over "simultaneous" equations system versus "recursive" equations systems that has prevailed for the last 30 years in the econometric literature and is still very live, by showing that the application of econometric estimation techniques to system dynamics models raises fundamental issues as to the "right" structure to represent the "real-world".

Mehra [24] states that there is very little advantage in engineering in using "simultaneous" equations systems, based on an analysis of the state-vector form of these models. Wold [91] has argued for several years that an appropriate formulation of economic models necessarily leads to recursive systems (since for Wold "the world is recursive")³⁶. Fisher's view [92] is that simultaneous models are limiting approximations to non-simultaneous models when certain time lags converge to zero. Senge [93] summarizes this approach very well and further elaborates on the issue in the light of the state-space approach...

Being unable to use 2SLS, the only other econometric method of interest available was the LISREL model³⁷

³⁶ This is also Forrester's view

³⁷ The work of Senge was extended to the investigation of the performance of IV estimators on the same model by J. Morecroft [94], but with only 5% measurement noise. The results, although better than the OLS ones, are far from

B. LISREL RUNS

The LISREL method, described in section III-D, was tried with several smaller scale models (as described in section (X-B)), as well as with the Market-Growth model, in order to investigate its behavior with the presence of lagged variables and specification of the order of nonlinearity found in the Market-Growth model.

B.1 The Identifiability Problem

The solution of the identifiability problem (see [64]) is often complicated and tedious; explicit solutions for all θ 's seldom exist. It is often difficult to determine whether or not a parameter is identified or whether or not the whole model is identified.

However, the program computes the information matrix at the starting point of the Fletcher-Powell search. If this matrix is positive definite, it is almost certain that the model is identified [64]. On the other hand, if it is singular, the model is not identified and an error message indicating a possible unidentified parameter in the parameter vector is printed out.

B.2 Experimental results

The program failed to converge, or displayed various singularity or other fatal errors, even with simpler models and since the actual listing was not available,

³⁷(cont'd)being satisfactory.

troubleshooting was very limited. It is a general weakness of "all-purpose" programs that they are hermetic to all except their "creator".

In view of their lack of interest and to keep this thesis concise, the results are not reported here.

The only alternative estimation technique open for investigation was the Filtering Form of the Maximum Likelihood algorithm, which is explained in detail in the next section.

IX. THE FILTERING FORM OF THE LIKELIHOOD FUNCTION

A. PHILOSOPHY OF THE MAXIMUM LIKELIHOOD METHOD

The Maximum Likelihood (ML) method was first introduced by R. A. Fisher in 1912 as a general method for statistical parameter estimation.

It can be explained as follows [22].

Let $(X_1, \dots, X_n)$ be a random sample, and let b be the object of estimation. Then, the joint density function of the sample evaluated at the sample values is:

$$f(X_1/b) \cdot \dots \cdot f(X_n/b) = L(X_1, \dots, X_n; b) \quad (1)$$

The product of the f 's on the left side specifies the probability density of obtaining the sample $(X_1, \dots, X_n)$ given that b is the parameter value (if the samples are independent). This product is written as $L(\cdot)$, which is known as the Likelihood function.

The main reason for the introduction of this function is that the expression in equation (1) can also be regarded as a function of b , given the sample, i.e., given that the sample has been drawn, the value of the likelihood depends on b .

The method of Maximum Likelihood proposes to use as an estimate of b , the function $b(X_1, \dots, X_n)$ which maximizes the Likelihood function:

$$L(X_1, \dots, X_n; b) \geq L(X_1, \dots, X_n; c) \quad (2)$$

where $c(X_1, \dots, X_n)$ is another estimator.

The underlying intuitive idea is that one should prefer a parameter value that implies a large probability of the sample drawn rather than a parameter value that declares that the sample is less probable.

B. MAXIMUM LIKELIHOOD AND IDENTIFICATION

" In practice, there are two major problems in obtaining Maximum Likelihood estimates for dynamic systems. These are:

1. Deriving an expression for the Likelihood function, and
2. Maximizing the Likelihood function with respect to the unknown parameters.

... The Likelihood function is the logarithm of the joint probability density of the observations given the parameters. If the observations are independent, the joint probability density function is easily written down since it is just the product of the probability densities of each observation given the parameters. The derivation of the Likelihood function becomes much more difficult when the observations are correlated. This is necessarily the case for dynamic systems with random inputs since the state at any time is correlated with the state at all the previous times".

However, it will be shown in section (IX-D) that the Likelihood function for a dynamic system can be derived in a

simple form using the Kalman Filter and the resulting white noise innovation sequence³⁸.

"... The second problem in obtaining Maximum Likelihood estimates is a computational one. Generally, the Likelihood function is highly nonlinear in terms of the parameters. For finite data lengths, it is also known to have several local maxima. In the case of dynamic systems, certain differential equation constraints have to be kept satisfied. The choice of a suitable search algorithm is very important for the successful application of Maximum Likelihood identification." [84]

The method originated in the work of Astrom [95], Kashyap [96] and Mehra [97] and is capable of solving the most general identification problem, including systems governed by nonlinear differential equations, the presence of additive white noise in the equations, and random disturbances corrupting measured variables.

The first application of the Maximum Likelihood method to system identification is due to Astrom and Bohlin [98] who considered single-input-single-output (SISO) and later multiple-input-single-output (MISO) systems of difference equation of the form:

³⁸ See section (IX-E) for description of the Kalman Filter and the definition of "innovation".

$$\begin{aligned}
& (1+a_1 \cdot q^{-1} + \dots + a_n \cdot q^{-n}) \bullet y(t) \\
& = \sum_{i=1}^p (b_{i1} \cdot q^{-1} + \dots + b_{in} \cdot q^{-n}) \bullet u_i(t) \\
& + l \bullet (1+c_1 \cdot q^{-1} + \dots + c_n \cdot q^{-n}) \bullet e(t)
\end{aligned} \tag{3}$$

or

$$A(q^{-1}) \bullet y(t) = \sum_{i=1}^p B_i(q^{-1}) \bullet u_i(t) + l \bullet C(q^{-1}) \bullet e(t) \tag{4}$$

where:

q : is the shift operator

p : the number of inputs

$e(t)$: is a sequence of independent gaussian noise

l : level of the noise signal

A more efficient computational procedure was derived by Eaton and presented two years later [99].

The generalization to identification using state-space models has been presented by Caines [100] and Woo [101], for example, at the second IFAC symposium on Identification and Process Parameter Estimation in 1970.

The linear form of these models is the classical:

$$\underline{x}(t+1) = A \bullet \underline{x}(t) + B \bullet \underline{u}t + G \bullet \underline{w}(t) \tag{5}$$

$$\underline{z}(t) = H \bullet \underline{x}(t) + \underline{v}(t) \tag{6}$$

where $\underline{x}$, $\underline{u}$, $\underline{z}$, $\underline{w}$, $\underline{v}$, are as defined in section (IX-E).

The purpose of identification is to find a model $M(b)$ that in some sense suitably describes the measured input-output data. Thus, the identification method depends largely on the criterion used to define "suitability".

The prediction of $\underline{z}(t+1)$ plays an important role for control. In, e.g Linear Quadratic Control Theory, the optimal input signal shall be chosen such that

$E\{\underline{z}(t+1)/Z(T)\}$ (where $Z(T) = \{\underline{z}(0) \dots \underline{z}(t)\}$) has desired behavior.

Therefore, it is very natural to choose a model that gives the best possible prediction.

That is, some function of the prediction error:

$$\underline{e}(t+1) = \underline{z}(t+1) - E\{\underline{z}(t+1)/Z(T), M(b)\} \quad (7)$$

should be minimized with respect to b .

Ljung [102] shows that prediction error criteria are intimately connected with the Maximum Likelihood approach. This is also done by Mehra [84] using the innovations approach³⁹.

The interest of the Maximum Likelihood approach lies in the fact that its formulation involves not only the prediction error terms, but also the covariance matrix of those terms, thus allowing us to use the maximum information available.

It is interesting that the Kalman Filter (to be discussed in section (IX-E)) yields, in a situation where both equation and measurement noise are present, the very terms that are needed for the computation of the Maximum Likelihood function, namely, the prediction error and its covariance matrix.

It was a natural step therefore to associate the Kalman Filter with the Maximum Likelihood algorithm. This coupling resulted in what has come to be known as the "Filtering Form of the Maximum Likelihood Algorithm".

³⁹ See section (IX-F)

By 1973, at the third IFAC Symposium in the Hague [84], participants were in the position to present numerous applications of this method to the identification of actual problematic cases and show the power of this algorithm to perform effectively in a situation where other estimators, and OLS in particular, yield biased estimates because, among other factors, of the measurement noise.

These include applications of the Maximum Likelihood algorithm to the identification of drum boiler dynamics [84, p. 87], aircraft parameters [84, p. 117], ship dynamics [84, p. 415], power generator dynamics [84, p. 367], nuclear power and reactor dynamics [84, p. 375], to quote only a few...

More recently, Ljung [102] solved the identifiability and consistency problems for the general linear, multi-dimensional time-invariant system under quite general assumptions, within the framework of prediction error identification methods. These results, as well as some other strong theorems about the multivariate case, have been used by e.g. Olsson [103] in further experiments in modeling and identification of a nuclear reactor.

An interesting development of the Maximum Likelihood algorithm is found in the work of Akaike [77], who recently motivated the use of Kullback's information criterion [74] in the context the Maximum Likelihood approach. He introduced an information criterion (AIC) for both estimation and statistical model determination defined by:

$$\text{AIC} = -2.(\text{maximum log-likelihood}) + 2.(\text{number of independently adjusted parameters}) \quad (8)$$

When several competing models are being fitted by the method of Maximum Likelihood, the one with the smallest number of parameters (the smallest value of AIC) is chosen as the best model. The final estimate is called the minimum AIC estimator (MAICE). This is a very elegant method for determining the structure of a representation for noisy input-output data and proves superior to traditional hypothesis testing methods.

" By the introduction of MAICE the problem of statistical identification is explicitly formulated as a problem of estimation and the need for subjective judgement required in the hypothesis testing procedure for the decision on the level of significance is completely eliminated." [78]

Rissanen [79] further expands on this theme and develops a new criterion for model-structure determination based on the concept of entropy.

The following sections give an intuitive description of this algorithm referred to in this thesis as LIKALM, for "LIKelihood using the KALMan filter", a brief summary of the Kalman Filter, and a formal derivation of the LIKALM algorithm. The method is first derived for a simple one-dimensional example. The general equations for the Kalman Filter and the LIKALM algorithm are then given.

C. AN INTUITIVE APPROACH TO THE FILTERING FORM OF THE ML ALGORITHM

The difference between the Least Squares and the LIKALM approach can be illustrated by a simple one dimensional example.

Let us assume the following model:

$$x(t+1) = b \bullet x(t) + u(t) \quad (9)$$

$$z(t+1) = x(t+1) + v(t+1) \quad (10)$$

where $u(t)$ and $v(t)$ are two independent gaussian sequences with known covariances:

$$E[u(t), u(t')] = Q \quad \text{iff } t = t' \quad (11)$$

$$E[v(t), v(t')] = R \quad \text{iff } t = t' \quad (12)$$

C.1 OLS

The OLS regression assumes the model:

$$z(t+1/t) = b \bullet z(t) \quad (13)$$

and chooses the estimate b of b such that the estimation criterion

$$S = \sum_{t=1}^T [z(t+1) - b \bullet z(t)]^2 \quad (14)$$

is minimized.

It is easy to show⁴⁰ that the value of b that minimizes this criterion is:

$$b = \frac{\sum_{t=1}^T [z(t+1) - z(t)]}{\sum_{t=1}^T z(t)^2} \quad (15)$$

⁴⁰See section (II-B.2)

the well-known Least Squares estimate.

This method is a "one shot" method in that it assumes all the data available and computes b by one pass through the data, since there is an analytic expression for b .

To make the comparison with LIKALM more obvious, we will assume here another interpretation of OLS.

Assuming that it were impossible to find a closed form solution to the minimization problem, we would have to use an iterative convergence algorithm to find b . Starting with a guessed value b_0 , we would simulate the model:

$$z(t+1/t) = b_0 \cdot z(t) \quad (16)$$

starting at the first data point; then compute the criterion:

$$S(z_t, b_0) = \sum_{t=1}^T [z(t+1/t) - z(t+1)]^2 \quad (17)$$

and compute the gradient:

$$\left. \frac{dS}{db} \right|_{b=b_0} = -2 \cdot \sum_{t=1}^T e(t) \cdot z(t) \quad (18)$$

If we are using the Newton-Raphson algorithm (for example), we would also compute the second derivative:

$$\left. \frac{d^2S}{db^2} \right|_{b=b_0} = -2 \cdot \sum_{t=1}^T z(t)^2 \quad (19)$$

We would then choose b_1 as:

$$b_1 = b_0 - \left[\frac{dS}{db} \cdot \left(\frac{d^2S}{db^2} \right)^{-1} \right] \bigg|_{b=b_0} \quad (20)$$

and reiterate the whole process until some convergence criterion is met.

The simulated trajectory for one iteration on b is shown in Figure IX-1. The trajectory is reinitialized at each data point.

This works well when there is no measurement noise. However, even small measurement errors lead to inaccurate estimates. This is best explained by Peterson [85].

"The intuitive reason for the failure is as follows. The motivation for reinitializing the model at each data point was to keep the simulation close to the true state of the system, so that any divergence between model and data, as measured by the residuals, would be meaningful. But reinitialization at noisy data is unlikely to keep the simulated state close to the real state. Large residuals may result from even a perfect model, under OLS."

C.2 LIKALM

The LIKALM approach differs from the above scheme in two points: the reinitialization of the simulation at each data point and the estimation criterion.

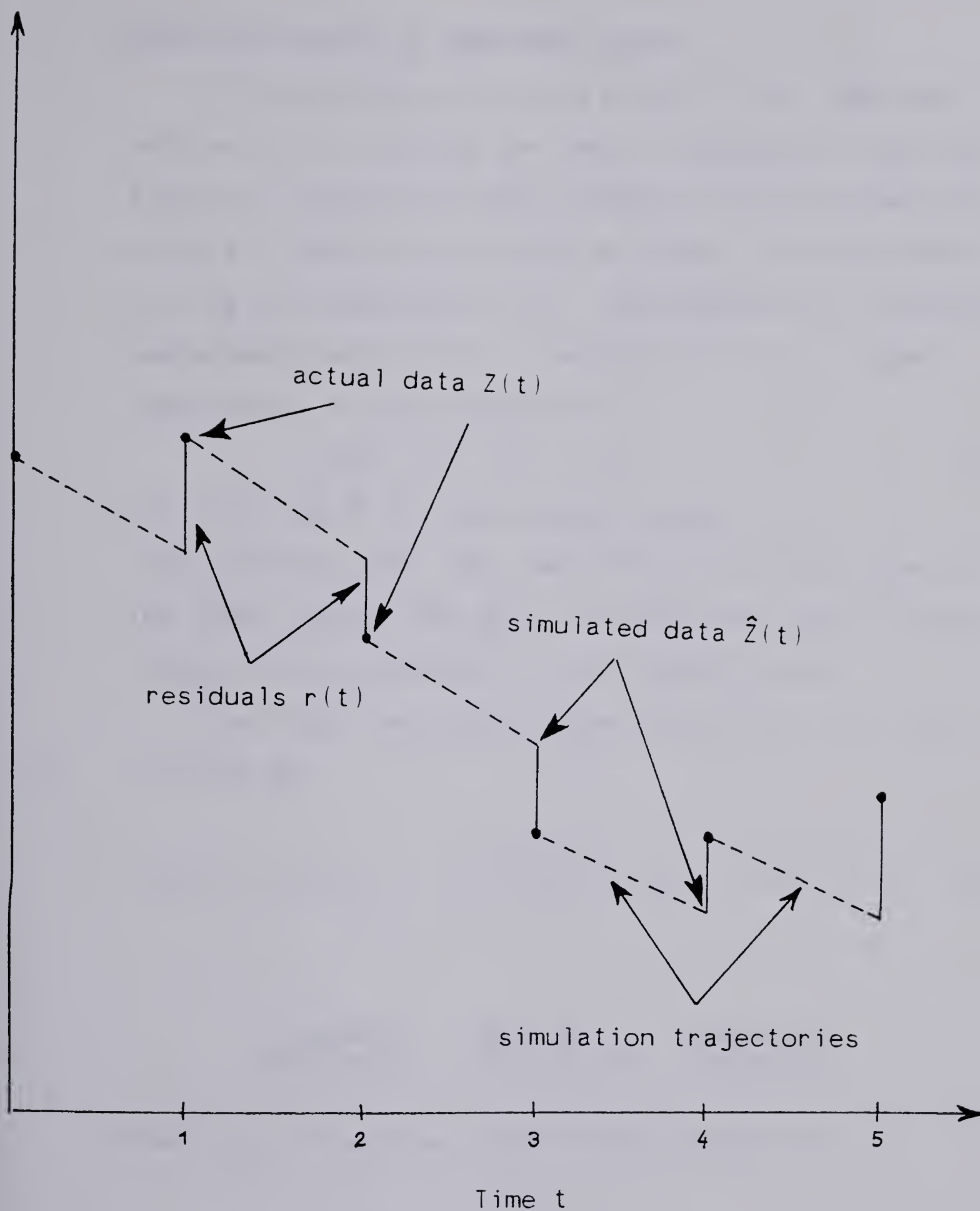


FIGURE IX-1: Reinitialization at each point in time in OLS simulation
(from [85])

Reinitialization at each data point

For each data point, the Kalman Filter (defined in section (IX-E)) yields the "best" estimate of the true state $x(t)$ as well as the variance of that estimate S_x , using all the data available up to and including time t . Calling this estimate $x(t/t)$, the prediction of the next measurement point $z(t+1)$, denoted $z(t+1/t)$, is then computed as in Least Squares by:

$$z(t+1/t) = b_i \cdot x(t/t) \quad (21)$$

for iteration i on the parameter value.

Thus, at each step, the simulation is reinitialized at the "best" point (the point in which the state is most likely to be, according to the Kalman Filter).

The "best" estimate of the state $x(t/t)$ at time t is given by:

$$x(t/t) = z(t/t-1) + \frac{S_x(t/t-1)}{S_z(t/t-1)} \cdot [z(t) - z(t/t-1)] \quad (22)$$

guessed measurement	gain or correction	error or innovation
------------------------	-----------------------	------------------------

where S_z is the covariance of the innovation.

The process is illustrated in Figure IX-2.

The estimation criterion

The Kalman filter yields not only the the residual or "innovation" $e(t) = z(t+1/t) - z(t+1)$ at each point,

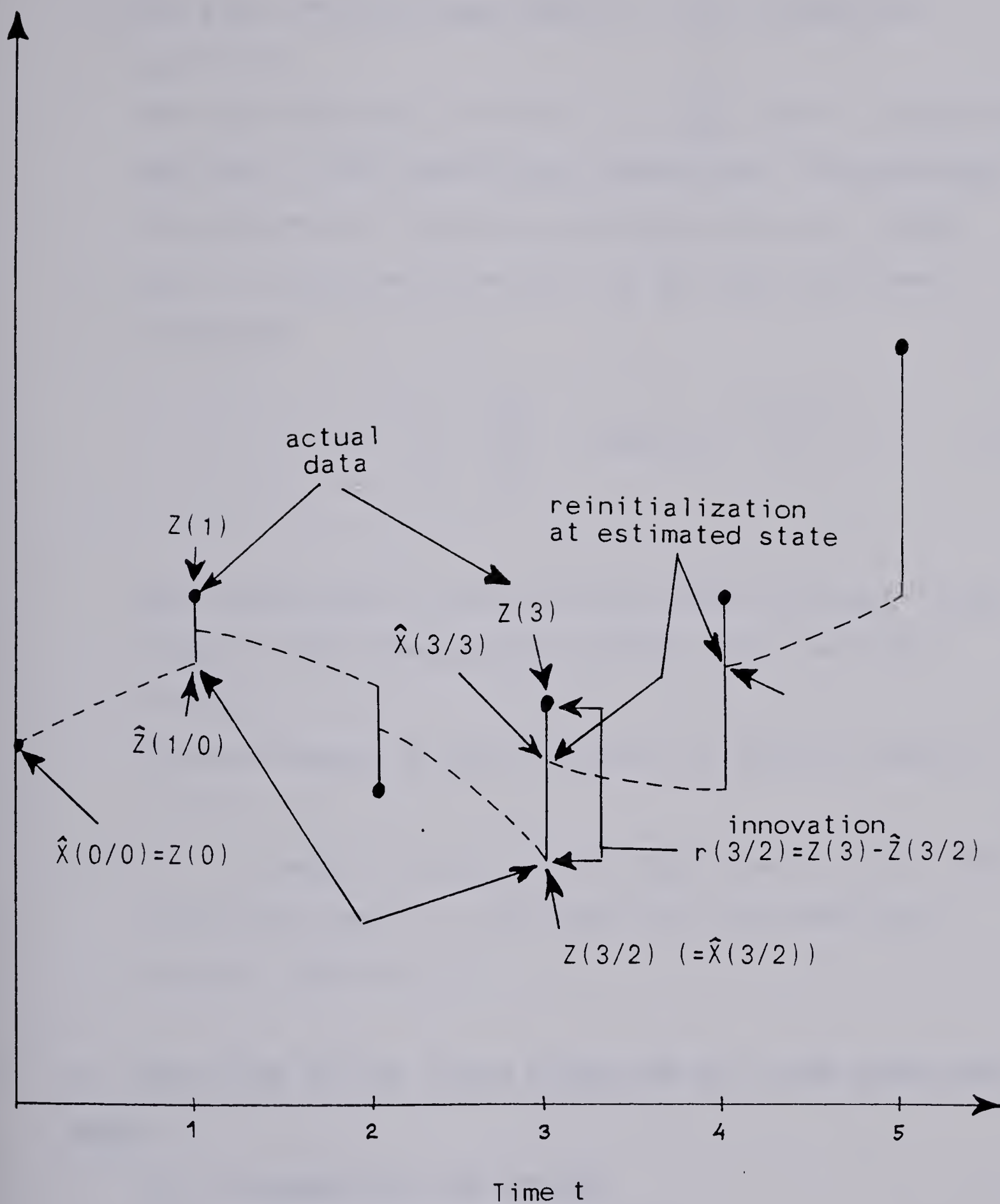


FIGURE IX-2: Reinitialization at each point in time in LIKALM simulation.
(from [85])

but also the covariance matrix of that innovation, $S_z(t+1/t)$.

The Least Squares criterion, $S = \sum_{t=1}^T [e(t)]^2$, would not make use of this additional information. Interestingly, the variance of the white noise process $e(t)$, noted $S_z(t)$, is precisely used by the Maximum Likelihood criterion:

$$L = K - \frac{1}{2} \sum_{t=1}^T \left[\text{Log}|S_z| + \frac{e^2(t)}{S} \right] \quad (23)$$

making the use of this criterion more efficient in the light of the information available from the Kalman Filter.

(The derivation of this criterion is left to section (IX-F).)

A simple illustration of these results with a one dimensional model is included with the experimental results. (Section (X-B.1))

D. DERIVATION OF THE LIKALM ALGORITHM FOR A ONE-DIMENSIONAL MODEL

D.1 Statement of the problem

Consider the following structure:

$$x(t+1) = b \cdot x(t) + u(t) \quad (24)$$

$$z(t+1) = x(t+1) + v(t+1) \quad (25)$$

with b unknown, $u(t)$ and $v(t)$ zero mean gaussian noise

sequences with known covariances:

$$E\{u(t), u(t')\} = s_u^2 = q \quad \text{iff } t = t' \quad (26)$$

$$E\{v(t), v(t')\} = s_v^2 = r \quad \text{iff } t = t' \quad (27)$$

We wish to analyze the performance of the scheme, for estimating b , that uses a Kalman Filter to estimate $x(T)$ and Maximum Likelihood, given $\{z(0), z(1), \dots, z(T)\}$

D.2 Solution of the problem

1. The Likelihood function

Given $\{z(0), z(1), \dots, z(T)\}$, and if we define this set of independent sample values up to time T as:

$Z(T) = \{z(0), z(1), \dots, z(T)\}$, we then have the following equalities for the Likelihood function, based on the assumption of normal distributions.

$$\begin{aligned} L\{[z(0), z(1), \dots, z(T)] | b\} &= p\{z(0), z(1), \dots, z(T) | b\} \\ &= p\{z(0) | b\} \cdot p\{z(1) | z(0), b\} \cdot \dots \cdot p\{z(T) | Z(T-1), b\} \\ &= p\{z(T) | Z(T-1), b\} \end{aligned} \quad (28)$$

so that L is now proportional to:

$$L = \prod_{t=0}^T \frac{1}{|s_z^2(t)|} \exp \left\{ - \frac{[z(t) - \hat{z}(t)]^2}{2 \cdot s_z^2} \right\} \quad (29)$$

where:

$$\hat{z}(t) = E\{z(t) | Z(t-1), b\} \quad (30)$$

$$s_z^2(t) = E\{[z(t) - \hat{z}(t)]^2 | Z(t-1), b\} \quad (31)$$

Taking the negative logarithm of L , we will minimize the criterion:

$$- \log L = \sum_{t=0}^T \left\{ \log |s_z(t)| + \frac{[z(t) - \hat{z}(t)]^2}{2 \cdot s_z^2(t)} \right\} \quad (32)$$

with respect to b .

2. The Kalman Filter

The ideas of Kalman Filtering can be put in the following simple form.

Supposing b known, we wish to estimate $x(t+1)$ from $Z(t) = \{z(0), z(1), \dots, z(t)\}$. To do this recursively, we suppose that $x(t)$ has previously been estimated from $Z(t-1)$. Then the two items of information available to us for the estimation of $x(t+1)$ are:

i) The "old" estimate:

$$\{x(t)|Z(t-1)\} = E\{x(t)|Z(t-1)\} \text{ or } \hat{x}(t/t-1) \quad (33)$$

ii) The "innovation"⁴¹ that is observed at time t :

$$e(t) = z(t) - \{z(t)|Z(t-1)\} = z(t) - \{x(t)|Z(t-1)\} \quad (34)$$

The last equality holds because $E\{v(t)\} = 0$.

It can be shown that these two items are statistically independent, using the conditional expectation theorem and the fact that they are Gaussian.

Consider now the two zero-mean Gaussian variates x and z defined as:

⁴¹ See p. 130

$$x = x(t+1) - E\{x(t+1)|Z(t-1)\} \quad (35)$$

$$z = e(t) \quad (36)$$

If z were not observed, then $E\{x\} = E\{x|Z(t-1)\}$ would be zero. When z is observed, the extent to which $E\{x|z\} = E\{x|Z(t)\}$ differs from zero is the "correction" that must be added to $E\{x(t+1)|Z(t-1)\}$ to obtain $x(t+1)$, in the light of the new information.

To obtain an expression for $E\{x|z\}$, let us consider the joint density:

$$p(x,z) = \frac{1}{2\pi \cdot s_x \cdot s_z \cdot \sqrt{1 - \rho^2}} \cdot \exp [A] \quad (37)$$

where:

$$A = \left\{ - \frac{1}{2(1 - \rho^2)} \left[\frac{x^2}{s_x^2} - \frac{2\rho xz}{s_x \cdot s_z} + \frac{z^2}{s_z^2} \right] \right\} \quad (38)$$

and s_x^2 and s_z^2 are variances, $\rho = s_{xz} / (s_x \cdot s_z)$ is the coefficient of correlation.

The conditional density is then:

$$p(x|z) = \frac{p(x,z)}{p(z)} = \frac{1}{\sqrt{2\pi} s_x \sqrt{1 - \rho^2}} \cdot \exp [B] \quad (39)$$

where:

$$B = - \frac{1}{2(1 - \rho^2)} \left\{ \frac{x^2}{s_x^2} - \frac{2\rho xz}{s_x \cdot s_z} + \frac{z^2}{s_z^2} - (1 - \rho^2) \frac{z^2}{s_z^2} \right\} \quad (40)$$

where the quantity in braces is a perfect square:

$$\{.\} = \frac{1}{s_x^2} (x - \rho \frac{s_x}{s_z} z)^2 \quad (41)$$

Then:

$$\begin{aligned} E\{x|z\} &= \int x \cdot p(x|z) \cdot dx \\ &= \int [x - \rho \cdot (s_x/s_z) \cdot z] \cdot p(x|z) \cdot dx \\ &\quad + \rho \cdot (s_x/s_z) \cdot z \cdot \int p(x|z) \cdot dx \\ &= \rho \cdot (s_x/s_z) \cdot z = (s_{xz} / s_z^2) \cdot z \end{aligned} \quad (42)$$

Thus the optimum estimate obeys:

$$\begin{aligned} x(t+1) &= E\{x(t+1) | Z(t)\} \\ &= E\{x(t+1) | Z(t-1)\} + (s_{xz} / s_z^2) \cdot e(t) \\ &= b \cdot x(t) + (s_{xz} / s_z^2) \cdot e(t) \end{aligned} \quad (43)$$

where:

$$\begin{aligned} s_z^2 &= E\{e^2(t)\} = E\{[x(t) - \hat{x}(t) + v(t)]^2\} \\ &= s_x^2(t) + r \end{aligned} \quad (44)$$

and

$$\begin{aligned} s_{xz} &= E\{[x(t+1) - E\{x(t+1) | Z(t-1)\}] \cdot [x(t) - \hat{x}(t) + v(t)]\} \\ &= E\{[b \cdot x(t) + u(t) - b \cdot x(t)] \cdot [x(t) - \hat{x}(t) + v(t)]\} \\ &= b \cdot s_x^2(t) \end{aligned} \quad (45)$$

Thus the Kalman Filter recursive equation is in this simple case:

$$x(t+1) = b \cdot x(t) + b \cdot \frac{s_x^2(t)}{s_x^2(t) + r} [z(t) - x(t)] \quad (46)$$

We also need a similar recursive equation for $s_x^2(t+1)$, which is given by:

$$\begin{aligned}
 s_x^2(t+1) &= E\{ [x(t+1) - \hat{x}(t+1)]^2 \} \\
 &= E\{ [b \cdot x(t) + u(t) - b \cdot \hat{x}(t) - (s_{xz}/s_z^2) \cdot e(t)]^2 \} \\
 &= b^2 \cdot s_x^2(t) + q - (s_{xz}^2/s_z^4) \cdot s_z^2 \\
 &= b^2 \cdot s_x^2(t) + q - b^2 \cdot [s_x^4(t) / (s_x^2 + r)] \quad (47)
 \end{aligned}$$

or:

$$s_x^2(t+1) = b^2 \cdot \frac{s_x^2(t) \cdot r}{s_x^2(t) + r} + q \quad (48)$$

It can be easily shown that the solutions of the variance equation are uniformly bounded for all t and that the "loop gain" of the filter equation is less than one if the initial model is stable: therefore, the Kalman Filter is internally stable.

3. Minimization of the Likelihood function

Now, to minimize the negative Likelihood:

$$-\text{Log } L = \sum_{t=0}^T \left\{ \text{Log } |s_z(t)| + \frac{[z(t) - \hat{z}(t)]^2}{2 \cdot s_z^2(t)} \right\} \quad (49)$$

with respect to b , we set:

$$\frac{d}{db} \{ -\text{Log } L \} = 0 \quad (50)$$

which yields:

$$\sum_{t=0}^T \left[\frac{1}{s_z(t)} \cdot \frac{ds_z(t)}{db} - 2 \frac{[z(t) - \hat{z}(t)]}{s_z^2(t)} \cdot \frac{dz(t)}{db} - 2 \frac{[z(t) - \hat{z}(t)]^2}{2 s_z^3(t)} \cdot \frac{ds_z(t)}{db} \right] = 0 \quad (51)$$

or:

$$\sum_{t=0}^T \left[\frac{1}{s_z(t)} \left(1 - \frac{[z(t) - \hat{z}(t)]^2}{s_z^2(t)} \right) \cdot \frac{ds_z(t)}{db} - \frac{[z(t) - \hat{z}(t)]}{s_z^2(t)} \cdot \frac{dz(t)}{db} \right] = 0 \quad (52)$$

Let us define:

$$c(t) = 1 - \frac{[z(t) - \hat{z}(t)]^2}{s_z^2(t)} \quad (53)$$

Because of the asymptotic stability of the Kalman Filter equations, as T becomes large, $s_z(t)$ and $(ds_z(t)/db)$ will approach constant values (which depend on b) while the multipliers $c(t)$ constitute a zero mean sequence, hence the relative contribution of the term in $(ds_z(t)/db)$ will vanish as T tends to infinity. Thus it suffices to set:

$$\lim_{T \rightarrow \infty} \sum_{t=0}^T \frac{[z(t) - \hat{z}(t)]}{s_z^2(t)} \cdot \frac{dz(t)}{db} = 0 \quad (54)$$

D.3 Unbiasedness of the LIKALM estimate

The precise behavior of the solutions b of equation (54) will depend on sample sequences $Z(T)$, but it is possible to show that if the iterative scheme illustrated in Figure IX-3 converges to some b_{lim} , this estimate will be unbiased, that is:

$$E[b_{lim}] = b \quad (55)$$

This can be established by writing (for the iterative scheme, $z(t)$ and $s_z^2(t)$ are fixed):

$$\sum_{t=0}^T \frac{[z(t) - \hat{z}(t)]}{s_z^2(t)} \cdot \frac{dz(t)}{db} = \sum_{t=0}^T \frac{[z(t) - \hat{z}(t)]}{s_z^2(t)} \cdot$$

$$\left\{ \hat{z}(t-1) + \frac{s_z^2(t)}{s_z^2(t) + r} [z(t-1) - \hat{z}(t-1)] \right\} \quad (56)$$

which has the expected value zero if, and only if:

$$E\{ [z(t) - \hat{z}(t)] \cdot \hat{z}(t-1) \} = 0 \quad (57)$$

and:

$$E\{ [z(t) - \hat{z}(t)] \cdot (z(t-1) - \hat{z}(t-1)) \} = 0 \quad (58)$$

But these conditions hold when the $z(t)$ are generated by a Kalman Filter with $b = b$, because then, and only then, are

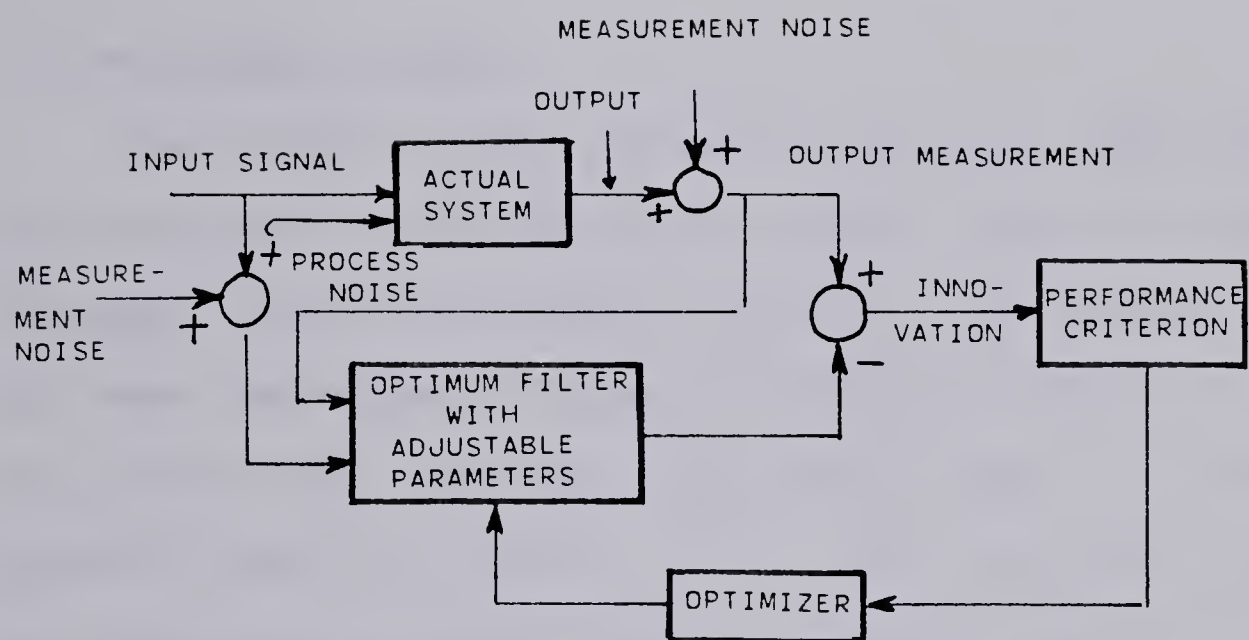


FIGURE IX-3: Iterative scheme for the LIKALM algorithm.
(from [24])

the serial correlations zero⁴².

This derivation and the properties of the Filtering Form of the Maximum Likelihood algorithm have of course been generalized to the multidimensional case. The next two sections present briefly the Kalman Filter and the derivation of the LIKALM algorithm for the multidimensional case.

E. THE KALMAN FILTER

The Kalman filter is one of the most important contributions to the field of control theory in the past two decades. Since its original derivation by Kalman [21], it has been rederived from many different points of view (see e.g. [68],[104]) and has been widely used in stochastic control (see e.g. [105], [106], [107]) and particularly in aerospace applications such as guidance and control. In its straightforward application, the Kalman filter involves the optimal estimation of the state variables of a system when the measurements are corrupted with white noise.

A brief summary of the Kalman filter, following Athans [27], is now presented.

E.1 Statement of the problem

Given a system in state-vector form:

$$\underline{x}(t) = F(t-1) \bullet \underline{x}(t-1) + B(t-1) \bullet \underline{u}(t-1) + G(t-1) \bullet \underline{w}(t-1) \quad (59)$$

where:

⁴²See comment on items i and ii on page 120

$\underline{x}(t)$ is the n-dimensional state vector

$\underline{u}(t)$ is an r-dimensional input vector

$\underline{w}(t)$ is a p-dimensional system noise vector

$F(t)$, $B(t)$, $G(t)$ are known matrices of appropriate order,

suppose the following measurement equation holds:

$$\underline{z}(t) = H(t) \bullet \underline{x}(t) + \underline{v}(t) \quad (60)$$

where:

$\underline{z}(t)$ is an m-dimensional output vector

$\underline{v}(t)$ is an m-dimensional vector of measurement noise

$H(t)$ is a known matrix,

and where the following assumptions hold:

1. $\underline{x}(0)$ is gaussian with known mean $\underline{x}(0)$ and known covariance matrix $S(0)$.

2. $\underline{w}(t)$ is gaussian with:

$$E[\underline{w}(t)] = 0 \text{ for all } t$$

$$\text{Cov}[\underline{w}(t), \underline{w}(t')] = Q(t) \text{ for } t = t' \text{ only}$$

$Q(t)$ is known positive definite.

3. $\underline{v}(t)$ is gaussian with:

$$E[\underline{v}(t)] = 0 \text{ for all } t$$

$$\text{Cov}[\underline{v}(t), \underline{v}(t')] = R \text{ for } t = t'$$

$R(t)$ is known positive definite

4. $\underline{x}(0)$, $\underline{w}(t)$ and $\underline{v}(t)$ are mutually independent for all values of t .

5. $\underline{u}(t)$ is a deterministic time sequence.

6. $F(t)$, $B(t)$, $G(t)$ and $H(t)$ are all deterministic.

It is desired to construct a "best" estimate of the state vector $\underline{x}(t)$ given past values of the input vector:

$$U(t-1) = [\underline{u}(0), \underline{u}(1), \dots, \underline{u}(t-1)]$$

and past values of the measurement vector:

$$Z(t) = [\underline{z}(1), \underline{z}(2), \dots, \underline{z}(t)]$$

The best estimate is denoted by $\underline{x}(t/t)$ and is defined as the conditional mean of $\underline{x}(t)$ given $U(t-1)$ and $Z(t)$, i.e.:

$$\underline{x}(t/t) = E[\underline{x}(t)/Z(t), U(t-1)].$$

E.2 Solution of the problem

The linearity of equations (59) and (60), together with the gaussian assumptions implies that the conditional probability density:

$$P[\underline{x}(t)/Z(t), U(t-1)] \quad (61)$$

is also gaussian, and hence, it is uniquely characterized by its conditional mean $\underline{x}(t/t)$ and conditional covariance matrix $S(t/t)$.

Kalman [21] derived a powerful sequential algorithm to generate $\underline{x}(t/t)$ and $S(t/t)$ using the properties of orthogonal projections (i.e. the error in the estimation of the state, $\underline{e} = \underline{x} - \underline{\hat{x}}$, is minimal if $\underline{\hat{x}}$ is the orthogonal projection of $\underline{x}$ into the space of "approximations")^{4 3}. We will only state the general result here.

1. The calculation of the conditional mean $\underline{x}(t/t)$ is done

^{4 3} The alternative derivation illustrated in section D-2 for a one-dimensional model can also be generalized

on-line using:

- initial condition:

$$\underline{x}(0/0) = \underline{x}(0) \quad (62)$$

- mean prediction:

$$\underline{x}(t/t-1) = F(t-1) \bullet \underline{x}(t-1/t-1) + B(t-1) \bullet \underline{u}(t-1) \quad (63)$$

- residual calculation:

$$\underline{e}(t) = \underline{z}(t) - H(t) \bullet \underline{x}(t/t-1) \quad (64)$$

- mean update equation:

$$\underline{x}(t/t) = \underline{x}(t/t-1) + K(t) \bullet \underline{e}(t) \quad (65)$$

where $K(t)$ can be computed from:

2. The calculation of the conditinal covariance matrix that can be done off-line using:

- initial value:

$$S(0/0) = S(0) \quad (66)$$

- covariance prediction equation:

$$\begin{aligned} S(t/t-1) = & F(t-1) \bullet S(t-1/t-1) \bullet F'(t-1) \\ & + G(t-1) \bullet Q(t-1) \bullet G'(t-1) \end{aligned} \quad (67)$$

- covariance update equation:

$$\begin{aligned} S(t/t) = & S(t/t-1) - S(t/t-1) \bullet H'(t) \bullet \\ & [H(t) \bullet S(t/t-1) \bullet H'(t) + R(t)]^{-1} \bullet \\ & H(t) \bullet S(t/t-1) \end{aligned} \quad (68)$$

- filter gain calculation:

$$K(t) = S(t/t) \bullet H'(t) \bullet R^{-1}(t) \quad (69)$$

The filter generates new state estimates as new observations become available, thus opening the possibility of real-time estimation. As an important and useful by-product, the filter generates the estimation

error covariance matrix, which measures the uncertainty in the estimate.

E.3 The Extended Kalman filter

The Kalman filter described above provides a very useful method of estimation of the states of linear systems. However the desire to use a method similar to the Kalman filter for nonlinear models has led to the formulation of the "extended" Kalman filter which is then only a suboptimal estimator (in the sense that an approximation is required to derive the estimator equations).

A nonlinear system (without external input and with additive noise, to simplify) can be written as:

$$\underline{x}(t) = \underline{f}(\underline{x}(t-1), t-1) + G(t-1) \bullet \underline{w}(t-1) \quad (70)$$

$$\underline{z}(t) = \underline{h}(\underline{x}(t), t) + \underline{v}(t) \quad (71)$$

where $\underline{x}$, $\underline{z}$, $\underline{w}$, $\underline{v}$, G are defined as previously and:

$\underline{f}(\underline{x}(t-1), t-1)$ is an n-dimensional nonlinear vector function of the state $\underline{x}(t)$ at time t

$\underline{h}(\underline{x}(t), t)$ is an m-dimensional nonlinear vector function of the state at time t .

all other assumptions remaining the same as in section (IX-E.1).

Since the solution becomes quickly computationally infeasible, even for very simple nonlinear systems, an "ad-hoc" approach is usually used to obtain practical equations.

Assuming that the "best" estimate of the state $\underline{x}(t-1/t-1)$ exists at time $t-1$, the idea is to expand

$\underline{f}(\underline{x}(t-1), t-1)$ about $\underline{x}(t-1/t-1)$ and $\underline{h}(\underline{x}(t), t)$ about $\underline{x}(t/t-1)$ in a Taylor series expansion and to retain only the first order terms.

As a result, the extended Kalman Filter equations can then be written:

$$\underline{x}(t/t-1) = \underline{f}(\underline{x}(t-1/t-1), t-1) \quad (72)$$

$$\underline{x}(t/t) = \underline{x}(t/t-1) + K(t) \bullet \underline{e}(t) \quad (73)$$

$$\underline{e}(t) = y(t) - \underline{h}(\underline{x}(t/t-1), t-1) \quad (74)$$

where $K(t)$ is calculated from the normal Kalman Filter equations with the matrices $H(t)$ and $F(t)$ defined by:

$$F(t) = d\underline{f}/d\underline{x} \text{ for } \underline{x} = \underline{x}(t/t) \quad (75)$$

$$H(t) = d\underline{h}/d\underline{x} \text{ for } \underline{x} = \underline{x}(t/t-1) \quad (76)$$

More advanced filters such as Second Order Filters, Single Stage Smoothing Filters, etc [108] can also be derived but this unnecessarily complicates an already complex situation and does probably not help much given the numerous uncertainties in the parameter estimation problem. More details on nonlinear filtering can be found in [68], [109], [110].

E.4 A word on innovations

The concept of "innovations" was introduced by Kailath [111]. In the equations of the Kalman Filter, the quantity:

$$\underline{e}(t) = z(t) - z(t/t-1) \quad (77)$$

is the new information brought about by the measurement $z(t)$. The sequence $\underline{e}(1), \dots, \underline{e}(t)$ is known as the "innovations sequence", and has been shown by Kailath [111]

to be zero mean, gaussian and white, and have the same statistics as the measurement noise process. These innovations represent the difference between the new measurement $z(t)$ and the prediction based on *all* the previous data $z(t/t-1)$. Thus the terms $e(t)$ contain the information contributed by the $z(t)$ so that $e(t)$ and $z(t)$ can be used interchangeably. One interest of introducing this concept is that the fact that the innovations sequence is gaussian allows us to formulate the Likelihood function in a simple way.

F. DERIVATION OF THE FILTERING FORM OF THE ML ESTIMATOR

This section follows closely Mehra [84]. It is assumed that the structure of the model is known. The vector of unknown parameters in F , B , G , H , Q and R and $\underline{x}(0)$ is denoted by $\underline{b}$.

It is assumed here that $\underline{b}$ is identifiable from a sequence of observations $\underline{z}(1) \dots \underline{z}(T)$ made on the system with noise and a sequence of known inputs.

Let $Z(T) = \{\underline{z}(1) \dots \underline{z}(T)\}$ be our sample of outputs.

The Likelihood function is then $L = P(Z(T)/\underline{b})$ and the maximum likelihood estimate is given by:

$$\underline{b} = \arg \{ \max_{\underline{b}} P(Z(T)/\underline{b}) \} \quad (78)$$

Using Bayes' Rule:

$$\begin{aligned} P(Z(T)/\underline{b}) &= P\{\underline{z}(1) \dots \underline{z}(T)/\underline{b}\} \\ &= P\{\underline{z}(T)/Z(T-1); \underline{b}\} \cdot P\{Z(T-1)/\underline{b}\} \end{aligned} \quad (79)$$

With successive applications of this rule, we get:

$$P\{Z(T)/\underline{b}\} = \prod_{t=1}^T P\{\underline{z}(t)/Z(t-1); \underline{b}\} \quad (80)$$

Since the Logarithm is a monotonic function, it is more convenient to write the Maximum Likelihood estimate as:

$$\underline{b} = \arg\{ \max/\underline{b} [\text{Log } P(Z(T)/\underline{b})] \} \quad (81)$$

$$= \arg \{ \max/\underline{b} \sum_{t=1}^T \text{Log } P[\underline{z}(t)/Z(t-1), \underline{b}] \} \quad (82)$$

If $\underline{x}(0)$, $\underline{w}(t)$, and $\underline{v}(t)$ are normally distributed, $P(\underline{z}(t)/Z(t-1), \underline{b})$ will also be normal and can be uniquely determined by computing the mean and covariance.

Therefore, let us define:

$$E\{\underline{z}(t)/Z(t-1), \underline{b}\} = \underline{z}(t/t-1) \quad (83)$$

and

$$\begin{aligned} \text{Cov}\{\underline{z}(t)/Z(t-1), \underline{b}\} &= E\{[\underline{z}(t) - \underline{z}(t/t-1)] \bullet [\underline{z}(t) - \underline{z}(t/t-1)]'\} \\ &= S_z(t/t-1) \end{aligned} \quad (84)$$

For a normally distributed k-vector $\underline{z}$, we have:

$$P(\underline{z}) = (1/[(2\pi)^{k/2} \bullet |S_z|^{1/2}]) \bullet \exp[(-1/2) \bullet \underline{e}' \bullet S_z^{-1} \bullet \underline{e}] \quad (85)$$

where $\underline{e} = \underline{z} - E(\underline{z})$ and $| |$ is the determinant.

Then:

$$\text{Log } P(\underline{z}) = \text{constant} - (1/2) \bullet \text{Log}|S_z| - (1/2) \bullet (\underline{e}' S_z^{-1} \underline{e}) \quad (86)$$

Therefore, our Log-likelihood function becomes:

$$\begin{aligned} L &= \text{constants} - (1/2) \bullet \sum_{t=1}^T \{ \text{Log } |S_z(t/t-1)| \\ &\quad + [(\underline{z}(t) - \underline{z}(t/t-1))]' \bullet S_z^{-1}(t/t-1) \\ &\quad \bullet [\underline{z}(t) - \underline{z}(t/t-1)] \} \end{aligned} \quad (87)$$

The problem of determining the Likelihood function has now become one of finding a way of calculating the conditional mean $\underline{z}(t/t-1)$ and the covariance matrix $S_z(t/t-1)$.

These quantities are precisely the output of a Kalman Filter given by (in a slightly different form than previously so that S_z may also appear:

Before measurement:

$$\underline{x}(t/t-1) = \underline{f}[\underline{x}(t-1/t-1), t-1] + B \bullet \underline{u}(t-1/t-1) \quad (88)$$

$$S_x(t/t-1) = F \bullet S_x(t-1/t-1) \bullet F' + GQG' \quad (89)$$

$$\underline{z}(t/t-1) = \underline{h}[\underline{x}(t/t-1), t] \quad (90)$$

$$S_z(t/t-1) = R + H \bullet S_x(t/t-1) \bullet H' \quad (91)$$

At time t we have the measurement $\underline{z}(t)$

Then the innovation is:

$$\underline{e}(t) = \underline{z}(t) - \underline{z}(t/t-1) \quad (92)$$

and the gain:

$$K(t) = S_x(t/t-1) \bullet H' \bullet S^{-1}(t/t-1) \quad (93)$$

And the update equations are:⁴⁴

$$\underline{x}(t/t) = \underline{x}(t/t-1) + K(t) \bullet \underline{e}(t) \quad (94)$$

$$S_x(t/t) = [I - K(t) \bullet H] \bullet S_x(t/t-1) \quad (95)$$

with: $S_x(0/0) = S_x(0)$ and $\underline{x}(0/0) = \underline{x}(0)$

The Maximum Likelihood estimate is now obtained by maximizing (87) with respect to $\underline{b}$ subject to the constraints in (88) to (91). This a very difficult optimization

⁴⁴Equ. (95) is an alternative formulation of equation (68).
 It follows from the matrix inversion lemma:
 If $A = B - BH'(HBH' + R)^{-1}HB$
 Then $A^{-1} = B^{-1} + H'R^{-1}H$
 See e.g. [104, p. 14]

problem...

In the case of linear systems, an approximation suggested by Eaton [99] simplifies the problem considerably. The gain K and the covariance matrix S_z are assumed to have attained constant values. The maximization of the loss function (87) can then be done analytically and in two steps:

Maximizing (87) with respect to S_z yields:

$$S_z = (1/T) \cdot \sum_{t=1}^T e(t/\underline{b}) \cdot e'(t/\underline{b}) \quad (96)$$

where $\underline{b}$ is given by the root of the equation:

$$\sum_{t=1}^T e'(t) \cdot S_z^{-1} \cdot \frac{de(t)}{d\underline{b}} = 0 \quad (97)$$

which is found by numerical minimization.

In the nonlinear case however, the usual approach is to treat the problem as a nonlinear programming problem in order to find an iterative solution to the parameter estimates.

Maximizing (87) is equivalent to minimizing the negative Likelihood function $J(\underline{b})$ with respect to $\underline{b}$.

The nonlinear programming algorithm is usually:

$$\underline{b}(i+1) = \underline{b}(i) - R_i \cdot \underline{g}_i \quad (98)$$

where:

$$\underline{g}_i = dJ/d\underline{b} \text{ for } \underline{b} = \underline{b}_i \quad (99)$$

R_i is an approximation to the second partial matrix, taken, in the Gauss-Newton method as the inverse of the information matrix M_i .

$$M_i = E[d^2J/d\underline{b}^2] \text{ for } \underline{b} = \underline{b}_i \quad (100)$$

F.1 Expressions of the gradient and the information matrix

Since the expression of J is:

$$J = -1/2 \sum_{t=1}^T \{ \underline{e}' \bullet S_z(t/t-1) \bullet \underline{e} + \text{Log}|S_z(t/t-1)| \} \quad (101)$$

differentiating with respect to b , we obtain the k -th component of the gradient as:

$$\begin{aligned} \left. \frac{dJ}{d\underline{b}} \right|_{\underline{b}=\underline{b}_i} &= \sum_{t=1}^T \left\{ \underline{e}' \bullet S_z^{-1} \bullet \frac{d\underline{e}}{db_k} - \frac{1}{2} (\underline{e}' \bullet \frac{dS_z}{db_k} \bullet S_z^{-1} \bullet \underline{e}) \right. \\ &\quad \left. + \frac{1}{2} \text{tr}(S_z^{-1} \bullet \frac{dS_z}{db_k}) \right\} \end{aligned} \quad (102)$$

where the time dependencies, $\underline{e}(t)$ and $S_z(t/t-1)$, have been neglected for the sake of clarity; and similarly the (k,l) element of the information matrix is given by:

$$\begin{aligned} M_{kl} &= \frac{d^2J}{db_k \bullet db_l} = \left\{ \frac{d\underline{e}}{db_k} \bullet S_z^{-1} \bullet \frac{d\underline{e}}{db_l} - \underline{e}' \bullet S_z^{-1} \bullet \frac{dS_z}{db_k} \bullet S_z^{-1} \bullet \frac{d\underline{e}}{db_l} \right. \\ &\quad \left. - \underline{e}' \bullet S_z^{-1} \bullet \frac{dS_z}{db_l} \bullet S_z^{-1} \bullet \frac{d\underline{e}}{db_k} - \frac{1}{2} \text{tr}[S_z^{-1} \bullet \frac{dS_z}{db_l} \bullet S_z^{-1} \bullet \frac{dS_z}{db_k}] \right\} \end{aligned} \quad (103)$$

neglecting the following 2nd order partials and first order partials of inverse matrices:

$$\begin{aligned}
 &+ (1/2) \operatorname{tr}[S_z^{-1} \bullet (d^2 S_z / db_k \bullet db_l)] \\
 &+ \underline{e}' \bullet S_z^{-1} \bullet (d\underline{e} / db_k \bullet db_l) \\
 &- (1/2) \underline{e}' \bullet S_z^{-1} \bullet (dS_z / db_l) \bullet S_z^{-1} \bullet (dS_z / db_k) \bullet S_z^{-1} \bullet \underline{e} \\
 &- (1/2) \underline{e}' \bullet S_z^{-1} \bullet (d^2 S_z / db_k \bullet db_l) \bullet S_z^{-1} \bullet \underline{e} \\
 &- (1/2) \underline{e}' \bullet S_z^{-1} \bullet (dS_z / db_k) \bullet S_z^{-1} \bullet (dS_z / db_l) \bullet \underline{e}
 \end{aligned}$$

In [112], Mehra proposes another approximation for the information matrix which turns out to be better in practice.

The above information matrix was obtained by neglecting a few second order terms in the expression:

$$M = d^2 J / d\underline{b}^2 \text{ and will be denoted } M_1.$$

The second form, that will be denoted M_2 here, is defined by:

$$\begin{aligned}
 M_2 &= E\{d^2 J / d\underline{b}^2\} \\
 &= E\{(dJ / d\underline{b}) \bullet (dJ / d\underline{b})'\} \quad (104)
 \end{aligned}$$

and is estimated from the sample as:

$$\begin{aligned}
 M(k, l) &= \sum_{k=1}^T \left\{ \left(\frac{d\underline{e}}{db(k)} \right)' \bullet S_z^{-1} \bullet \frac{d\underline{e}}{db(l)} \right. \\
 &\quad + \frac{1}{2} \operatorname{tr} \left[S_z^{-1} \bullet \frac{dS_z}{db(k)} \bullet S_z^{-1} \bullet \frac{dS_z}{db(l)} \right] \\
 &\quad \left. + \frac{1}{4} \operatorname{tr} \left\{ S_z^{-1} \bullet \frac{dS_z}{db(k)} \right\} \bullet \operatorname{tr} \left[S_z^{-1} \bullet \frac{dS_z}{db(l)} \right] \right\} \quad (105)
 \end{aligned}$$

This second version has the advantage of being always nonnegative definite so that one does not run into computational problems (such as negative diagonal elements) as do arise in the case of M_1 .

The flow chart for the Filtering Form of the Maximum Likelihood algorithm is then as illustrated in Figure IX-4.

F.2 Recursive equations

The gradient and the Information matrix involve the terms $d\bar{e}(t)/db(k)$ and $dS_z(t/t-1)/db(k)$. The set of recursive equations allowing to obtain these terms from a set of initial (zero) values are obtained by differentiating the Kalman Filter equations (88) to (95) and are called the "sensitivity equations" of the model. They can be obtained in a very straightforward manner from the Kalman Filter equations as follows:

Prediction:

$$\begin{aligned} d\bar{x}(t+1/t)/db(k) &= d\bar{f}(\bar{x}(t/t), t)/db(k) \\ &+ F(t) \bullet d\bar{x}(t/t)/db(k) \end{aligned} \quad (106)$$

$$\begin{aligned} dS_x(t+1/t) &= dF(t)/db(k) \bullet S_x(t/t) \bullet F'(t) \\ &+ F(t) \bullet dS_x(t/t)/db(k) \bullet F'(t) \\ &+ F(t) \bullet S_x(t/t) \bullet dF'(t)/db(k) \\ &+ dG(t)/db(k) \bullet Q(t) \bullet G'(t) \\ &+ G(t) \bullet dQ(t)/db(k) \bullet G'(t) \\ &+ G(t) \bullet Q(t) \bullet dG(t)/db(k) \end{aligned} \quad (107)$$

yielding:

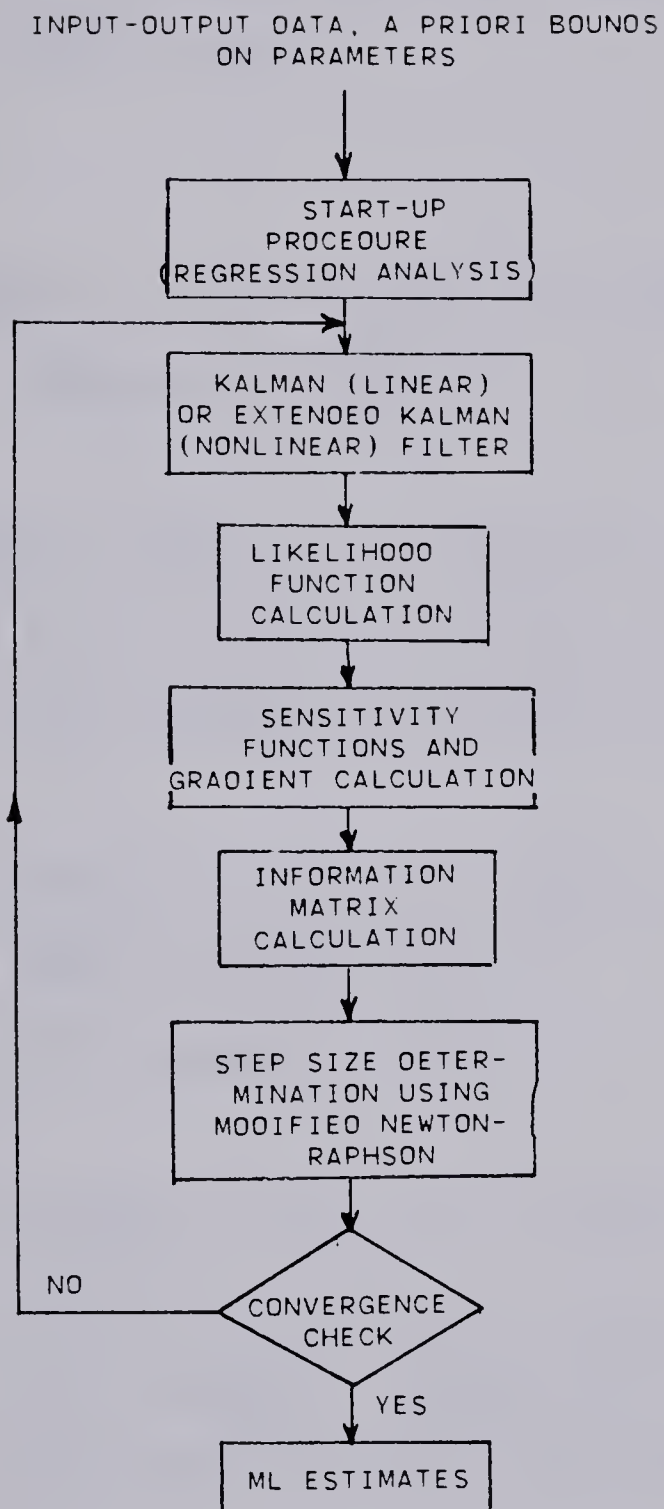


FIGURE IX-4: Flow-chart for the LIAKLM algorithm.
(from [24])

$$\begin{aligned}
dS_z(t+1/t)/d\underline{b}(k) &= [dH(t+1)/d\underline{b}(k)] \bullet S_x(t+1/t) \bullet H'(t+1) \\
&+ H(t+1) \bullet [dS_x(t+1/t)/d\underline{b}(k)] \bullet H'(t+1) \\
&+ H(t+1) \bullet S_x(t+1/t) \bullet [dH'(t+1)/d\underline{b}(k)] \\
&+ dR(t+1)/d\underline{b}(k)
\end{aligned} \tag{108}$$

Update:

$$\begin{aligned}
d\underline{e}(t+1)/d\underline{b}(k) &= - H(t+1) \bullet [d\underline{x}(t+1/t)/d\underline{b}(k)] \\
&- [d\underline{h}(\underline{x}(t+1/t)/d\underline{b}(k))]
\end{aligned} \tag{109}$$

$$\begin{aligned}
dK(t+1)/d\underline{b}(k) &= [dS_x(t+1/t)/d\underline{b}(k)] \bullet H'(t+1) \bullet S_z^{-1}(t+1/t) \\
&+ S_x(t+1/t) \bullet [dH'(t+1)/d\underline{b}(k)] \bullet S_z^{-1}(t+1/t) \\
&- K(t+1) \bullet [dS_z(t+1/t)/d\underline{b}(k)] \bullet S_z^{-1}(t+1/t)
\end{aligned} \tag{110}$$

$$\begin{aligned}
d\underline{x}(t+1/t+1)/d\underline{b}(k) &= [d\underline{x}(t+1/t)/d\underline{b}(k)] \\
&+ [dK(t+1)/d\underline{b}(k)] \bullet \underline{e}(t+1) \\
&- K(t+1) \bullet [d\underline{e}(t+1)/d\underline{b}(k)]
\end{aligned} \tag{111}$$

$$\begin{aligned}
dS_x(t+1/t+1)/d\underline{b}(k) &= [dS_x(t+1/t)/d\underline{b}(k)] \\
&- [dK(t+1)/d\underline{b}(k)] \bullet H(t+1) \bullet S_x(t+1/t) \\
&- K(t+1) \bullet [dH(t+1)/d\underline{b}(k)] \bullet S_x(t+1/t) \\
&- K(t+1) \bullet H(t+1) \bullet [dS_x(t+1/t)/d\underline{b}(k)]
\end{aligned} \tag{112}$$

with $dS_x(0/0) = 0$ and $d\underline{x}(0/0) = 0$.

X. ESTIMATION WITH THE FILTERING FORM OF THE MAXIMUM LIKELIHOOD

A. THE LIKALM COMPUTER PROGRAM

A computer program (LIKALM) using the Filtering Form of the Maximum Likelihood algorithm was written to estimate the model, since no such routine was available.

A.1 Main features of the LIKALM program

LIKALM is a general program for estimating any linear or nonlinear model written in state-space form (see equations in section (X-A.2)).

- A calling program reads in the options (explained later) and calls the appropriate subroutines.
- A subroutine FLET, called only if the minimization is done using the Fletcher-Powell algorithm instead of the information matrix. It reads in the initial values and parameters required by this procedure and calls subroutine FMFP (which in turn calls FUNCT as an external function). (see option 1 below).
- A subroutine FUNCT sets the fixed maximum dimensions of all arrays and calls the main subroutine FUNCT1. The maximum dimensions are presently set as follows:
 - dimension of the model (N): 10 state equations (or level variables)
 - number of parameters (L): 40 parameters estimated at one time (this includes parameters of the various covariance matrices involved, if they are estimated)

Two other dimensions provide additional flexibility: these are the dimension of the equation noise vector (M , maximum = 10) and the dimension of the output vector (or alternatively of the measurement noise vector)⁴⁵ (NZ , maximum = 10). In other words, not all the equations need contain noise and not all the states need to be measured.

- Subroutine FUNCT1 is the main program: it reads in all the dimensions and initial values from the input file (in variable format, the format for each input being provided by the user) and prints them out if required, reads in the data points from the data file (also in variable format), calls the Kalman Filter equations, the sensitivity equations, the computation of the gradient and the information matrix, the computation of the step size and checks for convergence⁴⁶

- Subroutines KALM, DIF, INFO, and STEP perform the computations called for by FUNCT1.

- Several algorithm and output options are provided:

1. An alternative minimization method is provided: this is the Fletcher Powell method [65]. The main difference between this method and the algorithm presented in section (IX-F) is that this method avoids the computation of the second order derivatives and performs a line search in the direction of steepest descent to determine the step size.

⁴⁵All measured outputs are assumed to be noisy

⁴⁶ No convergence check is yet provided: the program is stopped manually by the user. Setting a time limit on the run is a good idea

The public program FMFP from the IBM SSP library [113] performs this minimization. Subroutine FMFP requires the values of the function F to be minimized, and of its gradient, at different points. F and dF/dx are computed as external functions. In our case, F is the Likelihood function and dF/dx its gradient. Both are computed in subroutine INFO as in the normal algorithm.⁴⁷

2. The program can be run as an extended Kalman Filter only; in this case, the time t , and the values of the estimate of the state vector and its covariance matrix are printed.
3. A Least Squares subroutine OLSQ that operates on the values of the estimated state vector yielded by the Kalman filter is also provided if we wish to perform the iteration on the parameter vector using successive OLS estimates rather than the Maximum likelihood method⁴⁸.
4. The covariance matrices of the equation noise Q and the measurement noise R can be specified to be either both fixed or estimated or only one of them can be estimated by choosing the appropriate option.
5. Two output options allow the user to print out the headings (includes all the initial values) and the information matrix, its inverse the gradient vector and

⁴⁷The method does not seem to work properly however because the line search values affect the parameter vector so drastically that many of the matrices end up being ill conditioned

⁴⁸This method is not guaranteed to converge, but was tried as a simplification to LIKALM. The results were not convincing and are not presented here

the step vector. (The gradient and the step are always printed out at each iteration but in formats that can easily be subject to overflow if any problem arises)

■ In addition to the input file and the data file, the user has to write two FORTRAN subroutines STATE and DSTATE specifying the state equations and the matrix F (see section (IX-E.3)), and the derivatives of the state equations and F with respect to the parameter vector. If the Least Squares option is used, an OLS subroutine specifying the form of the estimated equations is also required.

A.2 The FORTRAN equations

The program assumes a data-generating state-space nonlinear model of the form^{4 9}

$$\underline{x}(t+1) = \text{FUN}[\underline{x}(t)] + G \bullet \underline{u}(t) \quad (1)$$

$$\underline{z}(t+1) = H \bullet \underline{x}(t+1) + \underline{v}(t+1) \quad (2)$$

with:

$$Q(t) = \text{cova}\{u(t)\} \quad (M \times M)$$

$$R(t) = \text{cova}\{v(t)\} \quad (NZ \times NZ)$$

$$SE0 = \text{cova}\{x(0)\} \quad (N \times N)$$

$$G: (N \times M)$$

$$H: (NZ \times N)$$

$$\underline{x}: (N \times 1)$$

$$\underline{z}: (NZ \times 1)$$

$$\underline{u}: (M \times 1) \text{ equation noise}$$

$$\underline{v}: (NZ \times 1) \text{ measurement noise}$$

^{4 9}No external input is provided since System dynamic models usually generate all dynamics within the boundaries of the system (see section (I-A.2)). This however could easily be added.

If we linearize the non-linear model about the nominal trajectory $x_{nom}(t) = \underline{x}(t/t)$ and $x_{nom}(t+1) = \underline{x}(t+1/t)$ (Taylor expansion limited to the first order terms), the equations of the Kalman Filter are⁵⁰:

if $F(\underline{x}(t/t)) = dFUN\{x(t)\}/dx$ at $x = \underline{x}(t/t)$

and adopting the conventions:

A = parameter vector $\underline{a}$ (LX1)⁵¹

$\underline{x}(t+1/t) = XP$ P for "predicted"

$\underline{x}(t/t) = XE$ E for "estimate"

$S_x(t+1/t) = SP$

$S_x(t/t) = SE$

$F[\underline{x}(t/t)] = F$

$S_z(t+1/t) = SZ$

$e(t+1) = NU$ ⁵²

$K(t+1) = KA$ ⁵³

$z(t+1) = Z$

$\underline{x}(t+1/t+1) = TXE$

$S_x(t+1/t+1) = TSE$

and

where T following a matrix denotes the transpose,

I following a matrix denotes the inverse of that matrix.

Then the equations of the Kalman filter become:

⁵⁰See section (IX-E)

⁵¹ The parameter vector and the two noise covariance matrices will be referred to as A , Q and R respectively, in all experimental results, even if they are only one-dimensional variables

⁵² The use of NU instead of e comes from an earlier notation for the innovations

⁵³To avoid the confusion with the index K in FORTRAN

Prediction:

$$XP = FUN(XE)$$

$$SP = F*SE*FT + G*Q*GT$$

$$SZ = H*SP*HT + R$$

Update:

$$NU = Z - H*XP$$

$$KA = SP*HT*SZI$$

$$TXE = XP + KA*NU$$

$$TSE = SP - KA*H*SP$$

and the sensitivity equations:

$$DXP = DFUN + F*DXE$$

$$DSP = DF*SE*FT + F*DSE*FT + F*SE*DFT$$

$$+ DG*Q*GT + G*DQ*GT + G*Q*DGT$$

$$DSZ = DH*SP*HT + H*DSP*HT + H*SP*DHT + DR$$

$$DNU = -H*DXP - DH*XP$$

$$DKA = DSP*HT*SZI + SP*DHT*SZI - KA*DSZ*SZI$$

$$DTXE = DXP + DKA*NU + KA*DNU$$

$$DTSE = DSP - DKA*H*SP - KA*DH*SP - KA*H*DSP$$

where $D(.)$ denotes $d(.) / dA$ and is a matrix of appropriate order and dimensions.

The negative Likelihood, its gradient and information matrix MA are then computed as:

$$J = 1/2 \sum (NUT*SZI*NU + LOG|SZ|)^{54} \quad (3)$$

$$DJ = \sum \{ (NUT*SZI*DNU - 1/2 NUT*SZI*DSZ*SZI*NU$$

⁵⁴ |.| denotes the determinant

$$+ 1/2 \text{ TR}(\text{SZI} * \text{DSZ})\}^{55} \quad (4)$$

$$\text{MA} = \sum \{(\text{DNUT} * \text{SZI} * \text{DNU} - \text{NUT} * \text{SZI} * \text{DSZ} * \text{SZI} * \text{DNU} \\ - \text{NUT} * \text{SZI} * \text{DSZ} * \text{SZI} * \text{DNU} - 1/2 \text{ TR}(\text{SZI} * \text{DSZ} * \text{SZI} * \text{DSZ})\} \quad (5)$$

Finally, the iteration step is computed as:

$$\text{DA} = \text{MAI} * \text{DJ} \quad (6)$$

and:

$$A_{i+1} = A_i - \alpha_i * \text{DA}_i \quad (7)$$

where α_i is a positive number adjusted to assure convergence (generally chosen equal to 1).

B. EXPERIMENTAL RESULTS

Four models ranging from simple linear to more complex nonlinear were chosen to test the performance of the estimation technique.

B.1 One Dimensional Model

The following results are meant as an illustration of the intuitive approach to LIKALM presented in section (IX-D). However, this will be more than just an illustration since it will allow us to explore the behavior of the Likelihood function, determine its minima manually and investigate the possibility of estimating the covariance matrices Q and R (which are simply variances in this case).

The Model

As already presented in section (IX-D.1) the model used here is:

$$x(t+1) = .8 \bullet x(t) + u(t) \quad (8)$$

⁵⁵ TR stands for trace

$$z(t+1) = x(t+1) + v(t+1) \quad (9)$$

x_0 is chosen as 10 and since the mean value of x , $\bar{x}$, is about 5, the standard deviations of $u(t)$ and $v(t)$ are chosen as 1% and 10% of $\bar{x}$ respectively, so that:

$$Q = 0.0025$$

$$R = 0.25$$

u and v are generated by the GGNOF routine⁵⁶ with seed numbers 1267 and 4328 respectively.

Three sets of experiments will be described here: the illustration of the LIKALM algorithm reinitialization procedure at each point in time, as opposed to that of OLS; the plotting of the Likelihood function and the manual determination of its minima with respect to different parameters; and the actual estimation of the model with a 100 data points:

1. Reinitialization in OLS and in LIKALM

To illustrate the reinitialization procedure explained in section (IX-C.1), the model was run with 10 data points, with $A_0 = .6$ (instead of the true value .8).

■ The Least Squares algorithm converges (with $dS/dA = 0$, where S is the criterion to be minimized) in two iterations yielding:

$$A = .74.$$

The two trajectories, with $A = .6$ and $A = .74$ are shown in Figures X-1 and X-2.

As the legend indicates, at each point in time, the

⁵⁶See p. 81

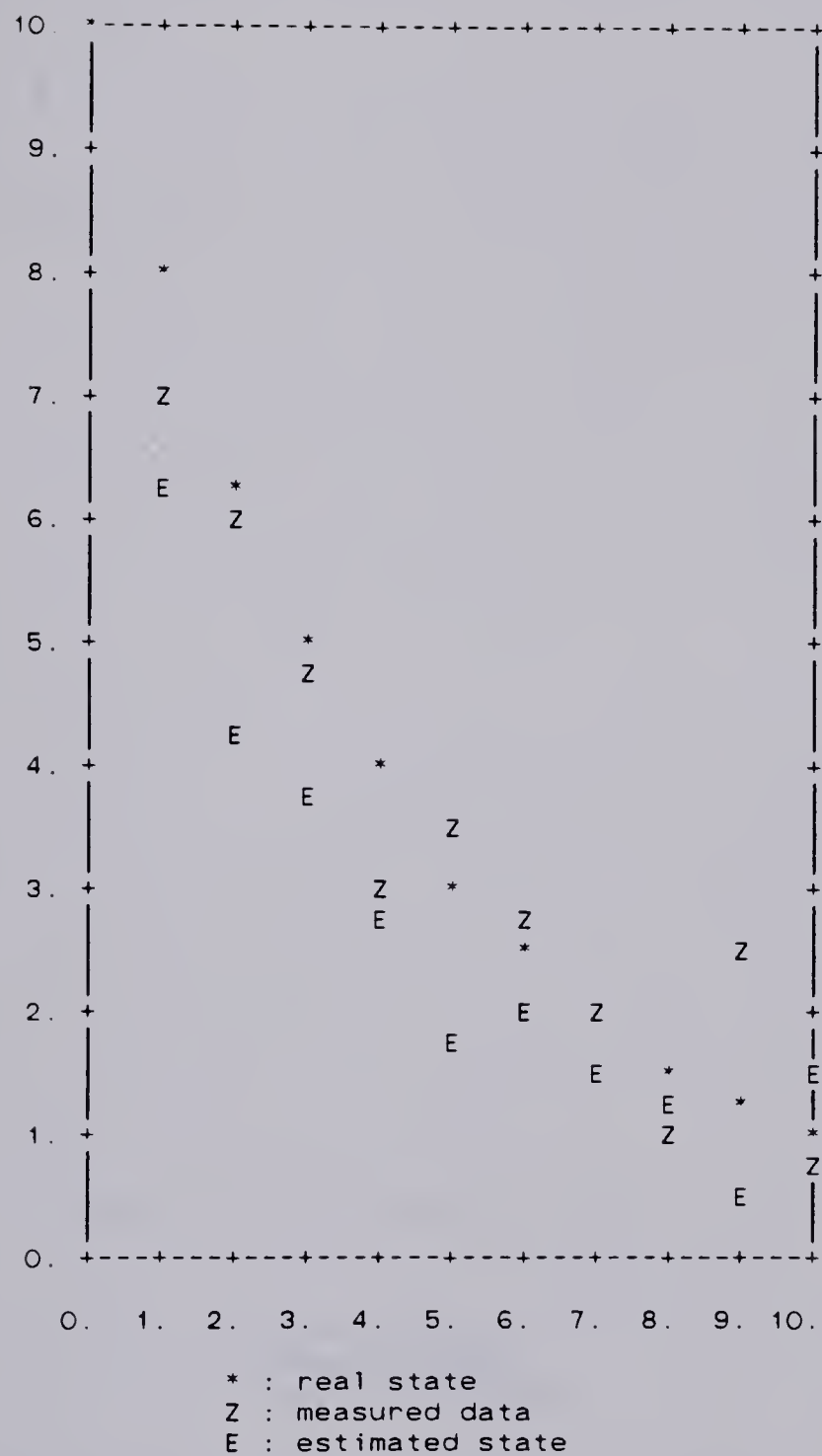


FIGURE X-1: One-dimensional model: simulation of trajectory using OLS (first iteration).

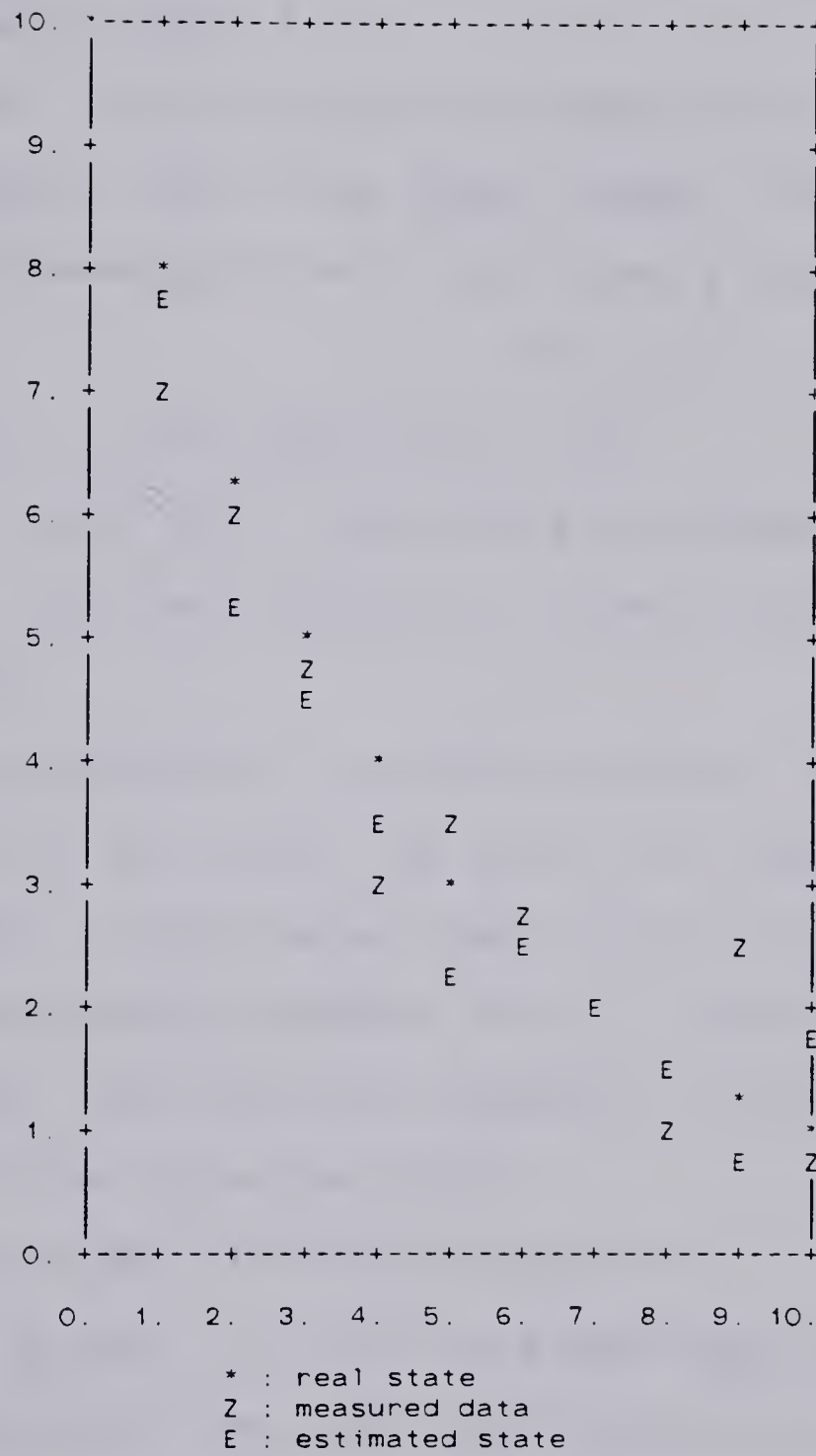


FIGURE X-2: One-dimensional model: simulation of trajectory using OLS (last (second) iteration).

simulation is reinitialized on the measured data point, thus keeping the trajectory far from the noiseless path. The values of x for the noiseless path, the noisy measurements z , and the estimated state $\hat{x}$ are also indicated in Table X-1.

■ The LIKALM program (run here with all unknowns specified at their true ideal values), starting with $A_0 = .6$ converges⁵⁷ in 4 iterations yielding:

$$A = .79 \pm 0.01$$

$$\text{with } L = 0.97 \text{ and } dL/dA = 3.27$$

The first and last trajectories are shown in Figures X-3 and X-4, and the values of x , z and $\hat{x}$ are shown in Table X-2.

Here the simulation is reinitialized at the Kalman estimate of the state. For the first iteration, the trajectory is much worse than for OLS since the Kalman Filter has wrong parameter values, but by the fourth iteration, the simulated trajectory is now remarkably closer to the noiseless path.

2. Plots of the Likelihood Function

It is easy, in this one dimensional case, to plot the behavior of the Likelihood function with respect to A , Q or R , to obtain an idea of its general shape, and find its minima manually in order to check whether the algorithm converges to the absolute minimum. Having obtained an intuitive grasp of the behavior of the

⁵⁷The convergence criterion, ϵ taken as the change in the likelihood function, was set to 0.001)

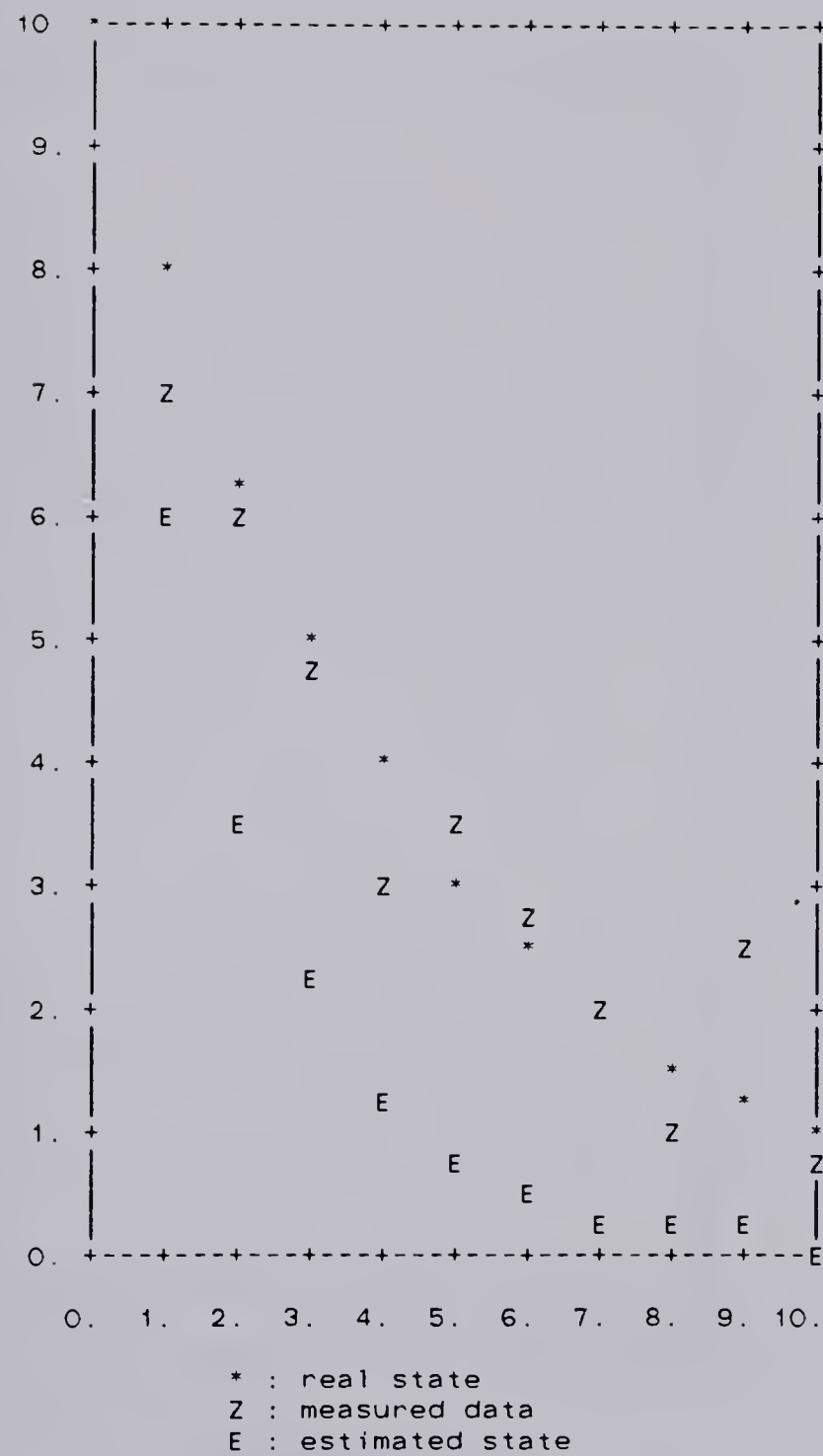


FIGURE X-3: One-dimensional model: simulation of trajectory using LIKALM (first iteration).

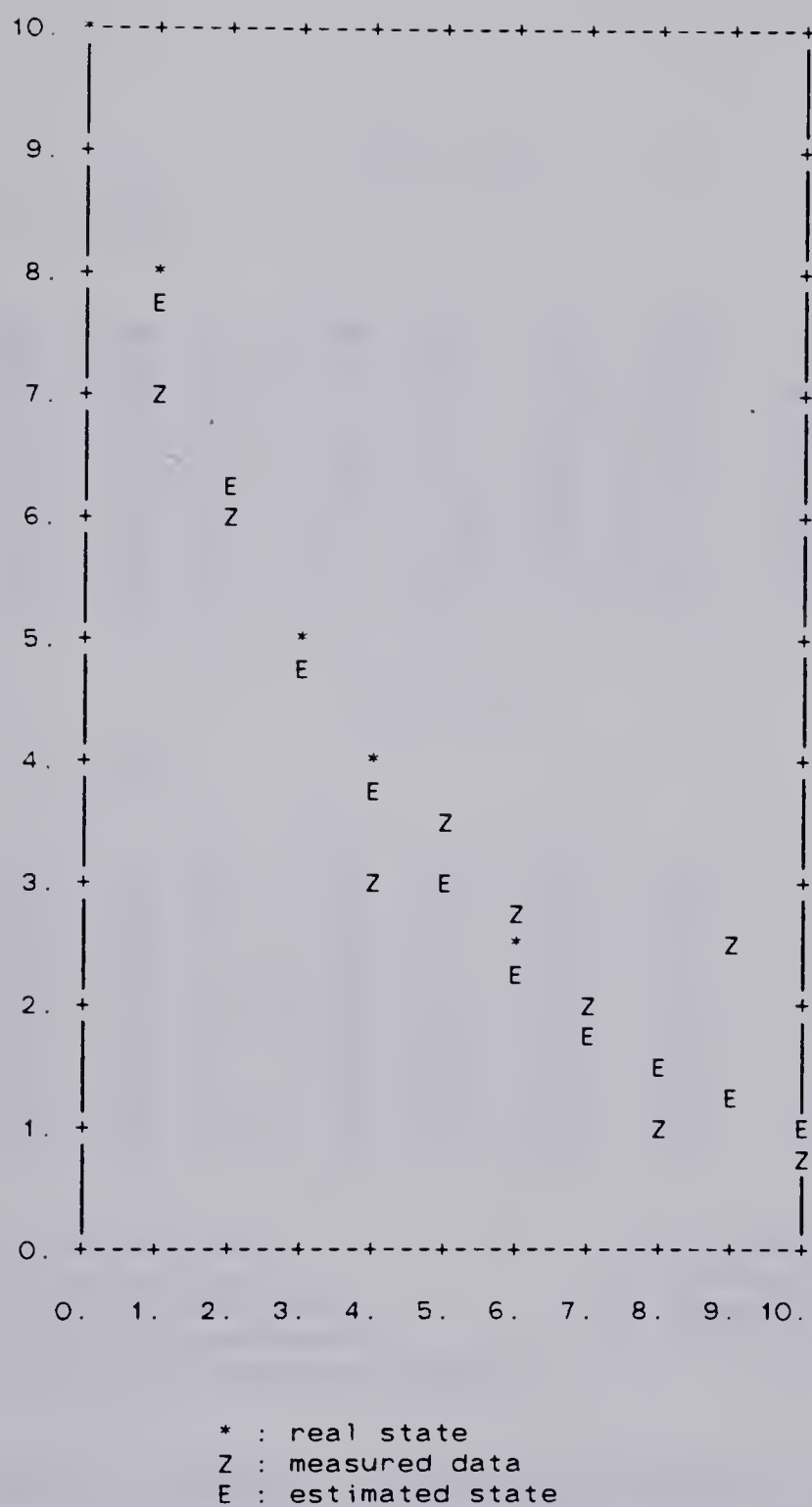


FIGURE X-4: One-dimensional model: simulation of trajectory using LIKALM (last (fourth) iteration).

ITERATION 1

A = 0.6000

t	X	TX	Z	ZP	E	XE	L	G	M
1	10.00	7.88	6.97	6.00	0.97	6.01	1.18	-5.40	11.76
2	7.88	6.33	6.06	3.61	2.45	3.64	12.36	-88.21	53.70
3	6.33	5.04	4.72	2.18	2.54	2.22	24.37	-250.11	405.70
4	5.04	3.97	3.08	1.33	1.75	1.36	29.73	-380.56	1215.73
5	3.97	3.07	3.52	0.82	2.71	0.86	43.48	-619.32	2072.87
6	3.07	2.45	2.65	0.51	2.13	0.55	51.76	-823.50	3407.00
7	2.45	1.97	2.11	0.33	1.78	0.35	57.32	-987.27	4764.79
8	1.97	1.52	0.95	0.21	0.73	0.22	57.70	-1040.47	5961.08
9	1.52	1.24	2.52	0.13	2.38	0.17	68.21	-1225.46	6634.08
10	1.24	0.96	0.76	0.10	0.66	0.11	68.39	-1269.65	7681.29

ITERATION 4

A = 0.7924

t	X	TX	Z	ZP	E	XE	L	G	M
1	10.00	7.88	6.97	7.86	-0.89	7.86	0.90	-4.29	11.76
2	7.88	6.33	6.06	6.18	-0.12	6.18	0.24	-2.61	71.70
3	6.33	5.04	4.72	4.86	-0.14	4.85	-0.41	-0.49	145.13
4	5.04	3.97	3.08	3.82	-0.73	3.80	-0.04	-13.54	232.16
5	3.97	3.07	3.52	2.99	0.53	3.00	-0.16	0.60	484.83
6	3.07	2.45	2.65	2.36	0.29	2.37	-0.69	5.68	559.89
7	2.45	1.97	2.11	1.86	0.25	1.87	-1.25	8.57	626.66
8	1.97	1.52	0.95	1.47	-0.52	1.46	-1.40	12.92	713.06
9	1.52	1.24	2.52	1.15	1.37	1.18	1.60	-6.29	795.58
10	1.24	0.96	0.76	0.93	-0.16	0.92	0.97	3.27	1137.57

X : state at time t	L : Likelihood function
TX : state at time t+1	G : gradient of L
Z : measured data at t+1	M : second derivative of S
ZP : predicted measurement	E : innovation
XE : estimate of state at time t+1	

TABLE X-2: Values of variables for figures X-3 and X-4.

algorithm under these circumstances, it becomes easier to understand its behavior when the multidimensionality makes it difficult to visualize.

Table X-3 presents the minima of the Likelihood function and the values of its gradient and second derivatives under various conditions.

In the case of $L(A)$, the initial value of the Likelihood function, $L = 68.39$ with a gradient of -479.8 is decreased to $L = 0.97$ with a gradient of 4.8 . Both versions of the Information matrix⁵⁸ $M1$ and $M2$ have similar positive values indicating that we do indeed have a minimum. In every case, the gradient sign change from $-$ to $+$ corresponds to the minimum of the Likelihood function.

It is interesting to note, however, that in the case of $L(Q)$ and $L(R)$, $M1$ behaves inconsistently since it is negative at the minimum! This confirms Mehra's statement⁵⁹ that this matrix can be non-positive definite and that it is preferable to use the second form, $M2$, which remains always positive.

Table X-3 also shows that the minimum $L(Q)$ is obtained for Q negative! This suggests that we probably are in presence of an identification problem since the Likelihood function does not seem to be able to display a minimum for the correct value of Q . The minimum for R

⁵⁸For a description of the two versions, see section (IX-F.1)

⁵⁹See section (IX-F.1)

TABLE X-3
 One-dimensional model
 Manual Minima of the Likelihood function

Conditions	Variable	Minimum	Likelihood	Information matrix	
				M1	M2
SS = 0.01 Q = 0.0025 R = 0.25	A	0.78	1.05	8671	8657
SS = 0.001 A = 0.8 R = 0.25	Q	0.045	1.71	-70	114
SS = 0.0001 A = 0.786 R = 0.25	Q	-0.084	0.95	-	2346
SS = 0.05 A = 0.8 Q = 0.0025	R	0.40	0.97	-30	15
SS = 0.05 A = 0.786 Q = 0.0025	R	0.35	0.38	-	19
SS = 0.05 A = 0.786 Q = 0.0045	R	0.35	0.77	-	15

Conditions:

10 data points

Q=1%, R=10%

SS is the step size for the manual search

is acceptable.

The plots of $L(A)$, $L(Q)$ and $L(R)$ in Figure X-5 show that the minimum for $L(A)$ is relatively well defined and attained via a steep slope. (The gradient varies from -383 for $A = 0.71$ to +2446 for $A = 0.9$). This is also true for $L(R)$ (with $A = A_{\min}$) where, dL/dR varies from -618 for $R = 0.05$ to 0 for $R = 0.35$, although the ascent after the minimum is much slower (the slope is only +3 for $R = 1$).

However, the situation is very different for Q for which $L(Q)$ (with $A = A_{\min}$) is extremely flat. L varies from 0.95 to 1.08 for Q varying from -0.01 to 0.03, with a corresponding variation for the gradient from -1.7 to 4.56, thus making it very difficult for the algorithm to locate the minimum.

Running the simulation with a hundred data points does not change this situation much.

If we now increase the equation noise level from 1% to 10% (R remaining at its previous 10% value), the Likelihood function $L(Q)$ displays a minimum for Q between 0.20 and 0.22, as indicated in Table X-4. The behavior of the Likelihood function with respect to Q and R is now very similar. They both attain their minimum in very close proximity to the true values, with similar Likelihood values at the minimum and a steep descent, as illustrated by Figure X-6.

The minimum for A is now much worse and the descent much

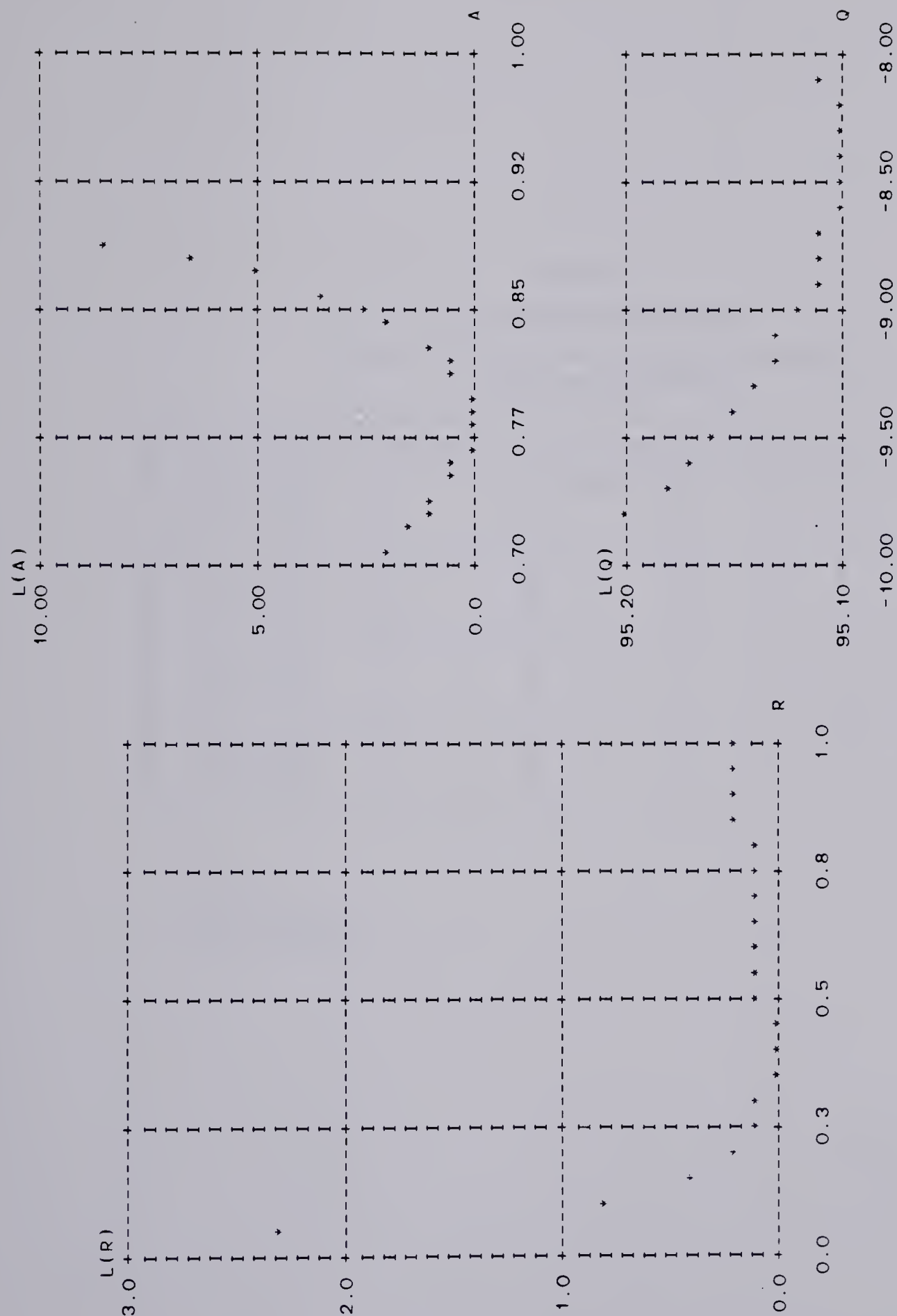


TABLE X-4

One-dimensional model

Manual Minima of the Likelihood function
10% equation noise

Conditions	Variable	Minimum	Likelihood	M
SS = 0.05 Q = 0.25 R = 0.25	A	0.68	17.96	707
SS = 0.05 A = 0.8 R = 0.25	Q	0.20	23.37	439
SS = 0.05 A = 0.68 R = 0.25	Q	0.20	17.68	429.3
SS = 0.05 A = 0.68 Q = 0.25	R	0.20	17.96	463.5

Conditions:

100 data points

Q=10%, R=10%

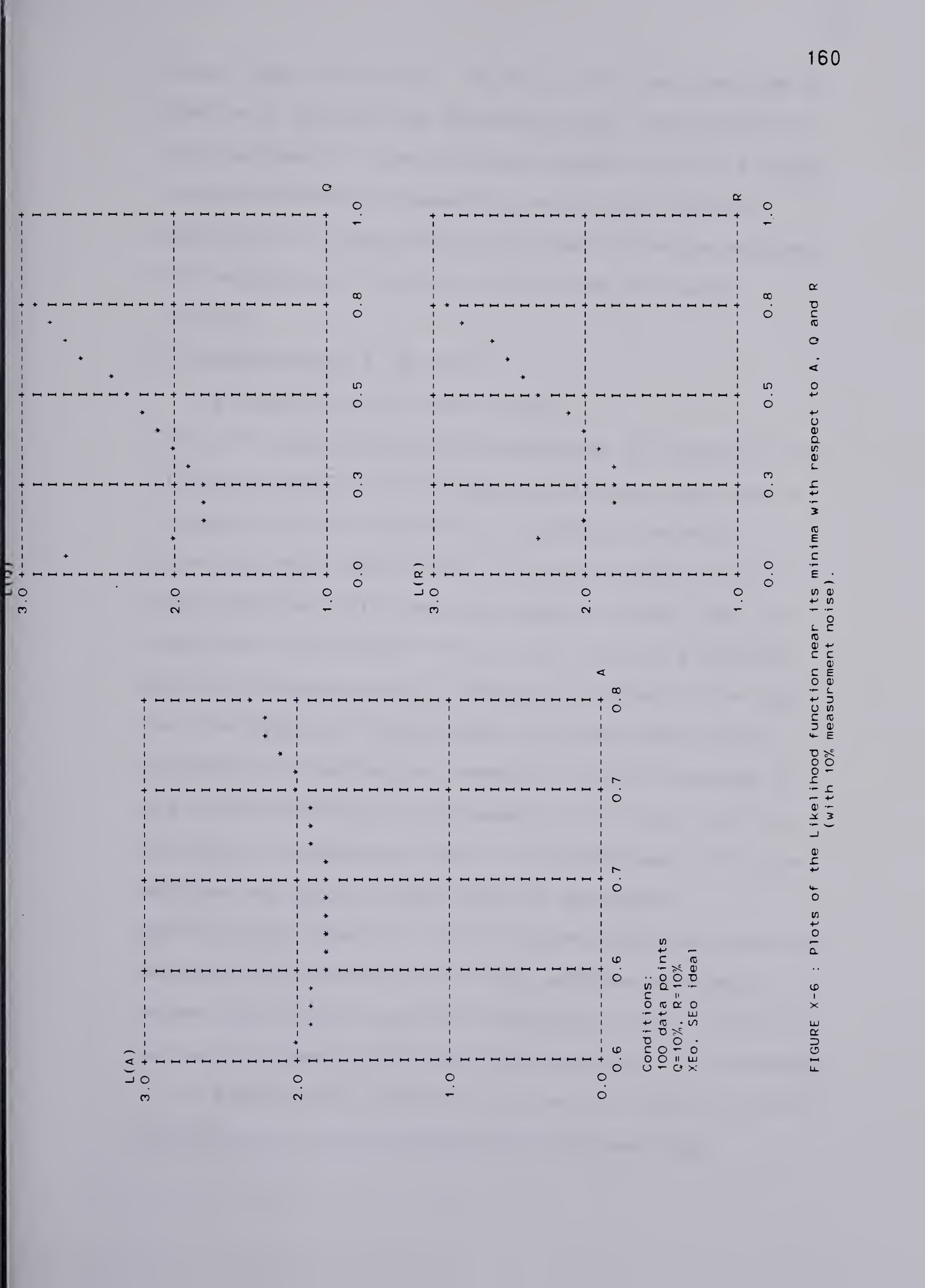


FIGURE X-6 : Plots of the Likelihood function near its minima with respect to A, Q and R (with 10% measurement noise).

slower than previously. The fact that the algorithm is capable of determining accurately only two out of the three parameters (the third one appearing with a large standard deviation) depending on the experimental conditions is indicative of an identification problem. This problem will be dealt with in the following sections.

3. Estimation of A , Q and R .

■ Description of the problem

One of the major problems encountered in the use of the Filtering Form of the Maximum Likelihood algorithm is the need to either specify or estimate the noise covariance matrices Q and R , as well as the initial state (X_{E0}) and its covariance matrix (S_{E0}) that are needed for the Kalman filter. This is also a problem generally encountered in the use of Kalman Filter when the statistics of the process and measurement noise sequences are not known. Generally, some knowledge of the noise statistics is assumed or available and the problem of estimating them is not even posed. In other engineering applications, such as aerospace applications, however, it is a rather difficult task to determine the statistics of many unknown sources of noise, and errors are then inevitable in the "a priori" values assigned to the covariance matrices. T. Nishimura [114] examines the effects of errors in these "a priori" statistics on the performance of the resulting

suboptimal filters. Although his study is limited to linear systems, it still shows that even in this case, the Kalman Filter behaves very poorly when the measurement noise covariance matrix R is either very large or very small compared to the true value.

Mehra [84] suggests using for R the sample covariance matrix obtained from the output error method⁶⁰. There is however no way of obtaining an estimate of Q , the process noise covariance matrix, unless some value is available from previous experiences, or the ratio of Q to R is known. Attempting to estimate both Q and R leads to identifiability problems, as illustrated below.

■ The identifiability problem

Referring back to section (IX-D), suppose that we now wish to estimate not only b , but also q and r . We can show with this simple model that there will be an identifiability problem.

Assuming that the Kalman Filter has attained a steady state, consider that the asymptotic algorithm ignores the gradient terms dsz/db , dsz/dq and dsz/dr in the Likelihood maximization and includes only the terms:

$$-\sum_{t=0}^T \frac{[z(t) - \underline{z}(t)]}{s^2} \bullet \text{grad}[\underline{z}(t, b, q, r)] \quad (10)$$

⁶⁰The output error method minimizes the square of the error between the actual system output and the output of a model used to represent the actual system. The method assumes measurement noise but no process noise

which is equivalent to minimization of the sample covariance of the innovations (assuming $sz(t)$ constant):

$$-\frac{1}{T} \sum_{t=0}^T [z(t) - \underline{z}(t, b, q, r)]^2 \quad (11)$$

But, as $T \rightarrow \infty$, this is asymptotic to:

$$\begin{aligned} E\{e\} &= E\{[x(t) - \underline{x}(t) + v(t)]^2\} = s_x^2 + r \\ &\geq s_x^2(\text{true}) + r(\text{true}) \end{aligned} \quad (12)$$

for a Kalman Filter with "true" parameters.

The algorithm will thus want to find values of b , q and r which minimize:

$$J = s_x^2 + r \quad (13)$$

subject to the Kalman Filter variance equation in its steady-state limit form:

$$s_x^2 = b^2 \cdot \frac{s_x^2 \cdot r}{s_x^2 + r} + q \quad (14)$$

But this problem is under-specified since, assuming a minimum value $J(\text{min})$ exists, with corresponding values of $s_x^2(\text{min})$ and $r(\text{min})$, any values of b and q satisfying the equation:

$$s_x^2(\text{min}) = b^2 \cdot \frac{s_x^2(\text{min}) \cdot r(\text{min})}{s_x^2(\text{min}) + r(\text{min})} + q \quad (15)$$

are solutions, hence are non-unique.

One can avoid this problem by fixing one parameter, say r , then minimizing s_x^2 as a function of b and q , and finally solving for a new r as found from the sample covariance matrix:

$$r = J(b, q, r) - s_x^2(b, q, r) \quad (16)$$

The iteration would be then repeated with r replacing r , and so on...

This, however, is applicable only in the case of linear systems, when the Kalman Filter can be assumed to have attained a steady-state, and is in essence the simplification proposed by Eaton [99] (see section (IX-F)). There is no theory as to what applies to the nonlinear case.

In this case, the general trend of practice seems to be to manipulate the information matrix by numerical methods when facing problems, as explained by Mehra [84], [112].

Stating that inaccurate noise statistics (usually chosen in an ad-hoc way or from "a priori" knowledge) are the cause of bias and identifiability problems in the use of the Extended Kalman Filter as a nonlinear state and parameter estimator, Ljung [115] proposes a different form of parameterization which could also be applied to the case of the LIKALM algorithm.

" In the model, there are certain assumptions associated with the noise covariance matrices,

whether parameterized or not. It should be noted that the effect of these assumptions is in fact only to provide a Kalman Filter gain. It is this gain that has the algorithmic importance and the noise assumptions are only vehicles to arrive at it. Therefore, it should in most cases be a good idea to parameterize the steady-state Kalman gain rather than the covariance matrices. This will normally involve fewer parameters (which means that usually the individual noise covariance are not identifiable - only the Kalman gain and the innovations covariance matrix are)." [115]

■ An "ad-hoc" solution

An ad-hoc way to circumvent the problem in our case is suggested here.

Referring back to section (II-C.4), in the case of a regression with a single variable, we have:

$$\underline{e} = \underline{y} - b \bullet \underline{x} \quad (17)$$

where $\underline{e}$, $\underline{y}$ and $\underline{x}$ are $(T \times 1)$ and T is the number of observations

An estimate of the residual sequence is obtained by replacing b by its estimate $\hat{b}$:

$$\underline{e} = \underline{y} - \hat{b} \bullet \underline{x}. \quad (18)$$

An estimate of the variance of the noise process is then given by:

$$s_e^2 = (1/T) \bullet \underline{e}' \underline{e}. \quad (19)$$

If we now have measurement noise added, adopting the same notations as in section (II-C.4), we have:

$$\underline{w} = \underline{e} + \underline{v} - b\underline{u} \quad (20)$$

where $\underline{w} = \underline{Y} - b \bullet \underline{X}$

and we have the following relationship between the various noise variances:

$$s_w^2 = s_e^2 + s_v^2 + b^2 \bullet s_u^2 \quad (21)$$

which can be estimated as:

$$s_w^2 = s_e^2 + s_v^2 + b^2 \bullet s_u^2 \quad (22)$$

In the multiple regression case, the situation becomes more complex.

If we now have:

$$\underline{e} = \underline{y} - X\underline{b} \quad (23)$$

where $\underline{e}$ and $\underline{y}$ are $(T \times 1)$, X is $(T \times N)$, $\underline{b}$ is $(N \times 1)$ and N is the number of regressors.

If $\underline{x}(1) \dots \underline{x}(N)$ are measured with noise such that $\underline{z}(1) = \underline{x}(1) + \underline{u}(1) \dots \underline{z}(N) = \underline{x}(N) + \underline{u}(N)$, and $\underline{Y} = \underline{y} + \underline{v}$, then:

$$\underline{Y} = Z\underline{b} + \underline{w} \quad (24)$$

where:

$$\underline{w} = \underline{e} + \underline{v} - U\underline{b} \quad (25)$$

and:

$$Z = \{\underline{u}(1) \dots \underline{u}(N)\}$$

$$U = \{\underline{u}(1) \dots \underline{u}(N)\}$$

We then obtain an expression for the variance of the

noise process as:

$$s_w^2 = s_e^2 + s_v^2 + \underline{b}U'U\underline{b} \quad (26)$$

assuming again that all noise sequences are not correlated.

If the system is in system dynamics formulation so that each equation can be written in the form:

$$\underline{Y} = YL' \bullet \underline{b} + \underline{e} \quad (27)$$

where YL is $(T \times N)$, matrix of lagged values of all endogenous variables, and if we measure all the variables with noise, equation (26) becomes:

$$s_w^2 = s_e^2 + s_u^2 + \underline{b}' U' U \underline{b} \quad (28)$$

But $U'U$ now becomes the measurement noise covariance matrix R (assumed diagonal).

Thus, if we know the ratio between equation and measurement noise for each equation, the above equation can be solved for r_{ii} and q_{ii} .

This again demonstrates that our problem is under-specified and therefore nonuniquely solvable. sw^2 can be computed from the sum of squared residuals SSR yielded by the TSP Least Squares routine [89]. A diagonal matrix S can thus be constructed from the SSR values for each individual equation and we have the relationship:

$$S = Q + R + M \quad (29)$$

where M is a matrix involving terms of the type $\underline{b}' U' U \underline{b}$ or $\underline{b}' R \underline{b}$.

The magnitude of the terms in M depends on the

estimates of the parameter values as well as on the various measurement noise variances. The terms in M are however always positive so that S is an upper bound on $Q + R$. In the absence of any information at all on the order of magnitude of the noise variances (which is a purely theoretical situation since in most practical cases the modeler would have at least some idea of the upper bounds of such errors), it is suggested here to use the matrix S to obtain values for R and Q based on the approximation $S = Q + R$ (very rough in the case of large parameter values).

One could then advance by trial and error in assessing an approximate value for the ratio, guided by the value of the Likelihood function. Again, it would help here to have an idea of the ratio Q/R ...

■ Experimental results.

Attempts at estimating various combinations of A , Q and R in the one-dimensional model can be summarized in the following qualitative way.

- If Q is estimated alone or along with A , the final value of Q is negative with a large standard deviation, or zero if Q is constrained to remain positive. The estimate for A is good.
- If R is estimated alone or with A , the final values for both parameters are acceptable.
- If Q and R are estimated alone, Q is still negative

and R is acceptable.

- If A , Q and R are estimated, the algorithm does not converge unless Q is constrained to remain positive.

Table X-5 presents a comparison of the estimates of A , Q and R with 1% and 10% equation noise (the measurement noise is kept at 10% in both cases).

In the $Q=1\%$ case, with Q constrained to remain positive, A and R are estimated very accurately and are within one standard deviation of their true values.

With $Q=10\%$, the estimate of Q is now much better, but the value and standard deviation of A have worsened, as expected from the results of the plots of the Likelihood function.

In the light of these results with a simple one-dimensional model, we can expect difficulties in the estimation of the of the noise statistics for higher order systems. This situation is further examined with our next model, a two-dimensional oscillator.

B.2 Two Dimensional Oscillator

1. The model

This is a very simple model destined to make sure that the algorithm behaves properly with a linear model, when there is interaction between equations, before attempting to try it with a more complex nonlinear model. Again, the simplicity of this model and its small size make it possible to examine thoroughly the behavior of the algorithm under different conditions.

Table X-5

One-dimensional model

Estimates of A, Q and R for
1% and 10% equation noise

Noise	Conditions	A	Q	R	Likelihood	iter.
Q = 0.0025 R = 0.25 A = .8	Ao = .6 Qo ideal Ro = 0.5 SEo = 0.5 XEo = Z1	0.8218 (0.0123)	0	0.2600 (0.0382)	-23.39	5
Q = 0.25 R = 0.25 A = .8	Ao = .6 Qo = 0.1 Ro = 0.1 SEo ideal XEo ideal	0.6923 (0.0342)	0.1709 (0.0713)	0.2894 (0.0783)	17.53	6

Conditions:

100 data points

The model is:

$$TX1 = X1 + 0.01 \cdot X2 + U1 \quad (30)$$

$$TX2 = X2 - 0.01 \cdot X1 + U2 \quad (31)$$

and:

$$Y1 = TX1 + V1 \quad (32)$$

$$Y2 = TX2 + V2 \quad (33)$$

Where U and V are random sequences generated by the IMSL subroutine GGNOF using different seeds and with standard deviations varying with experimental conditions.

The unknown parameters are $A1 = 0.01$, $A2 = -0.01$ and the variances of the noise sequences are equal to 1% and 10% of the mean value of $X1$ or $X2$: $\bar{x} = 1.5$, so that:

$$s_u = 0.015 \text{ and } Q = s_u^2 = 0.000225$$

$$s_v = 0.15 \text{ and } R = s_v^2 = 0.0225$$

The initial values for $X1$ and $X2$ are chosen as 2. and 0. respectively. The trajectory of the model with and without the noise inputs is illustrated in Figure X-7.

2. OLS estimates

The OLS estimates of this simple model with 1% equation noise and 10% measurement noise are presented in Table X-6, along with the seed numbers for GGNOF.

The estimate of $A1$ is very bad, being in error by a factor of about 3 and is also insignificant (probability that $A1 = 0$ is 85% !) The fit is very low, thus indicating the drastic effect of the noise on this simple linear model.

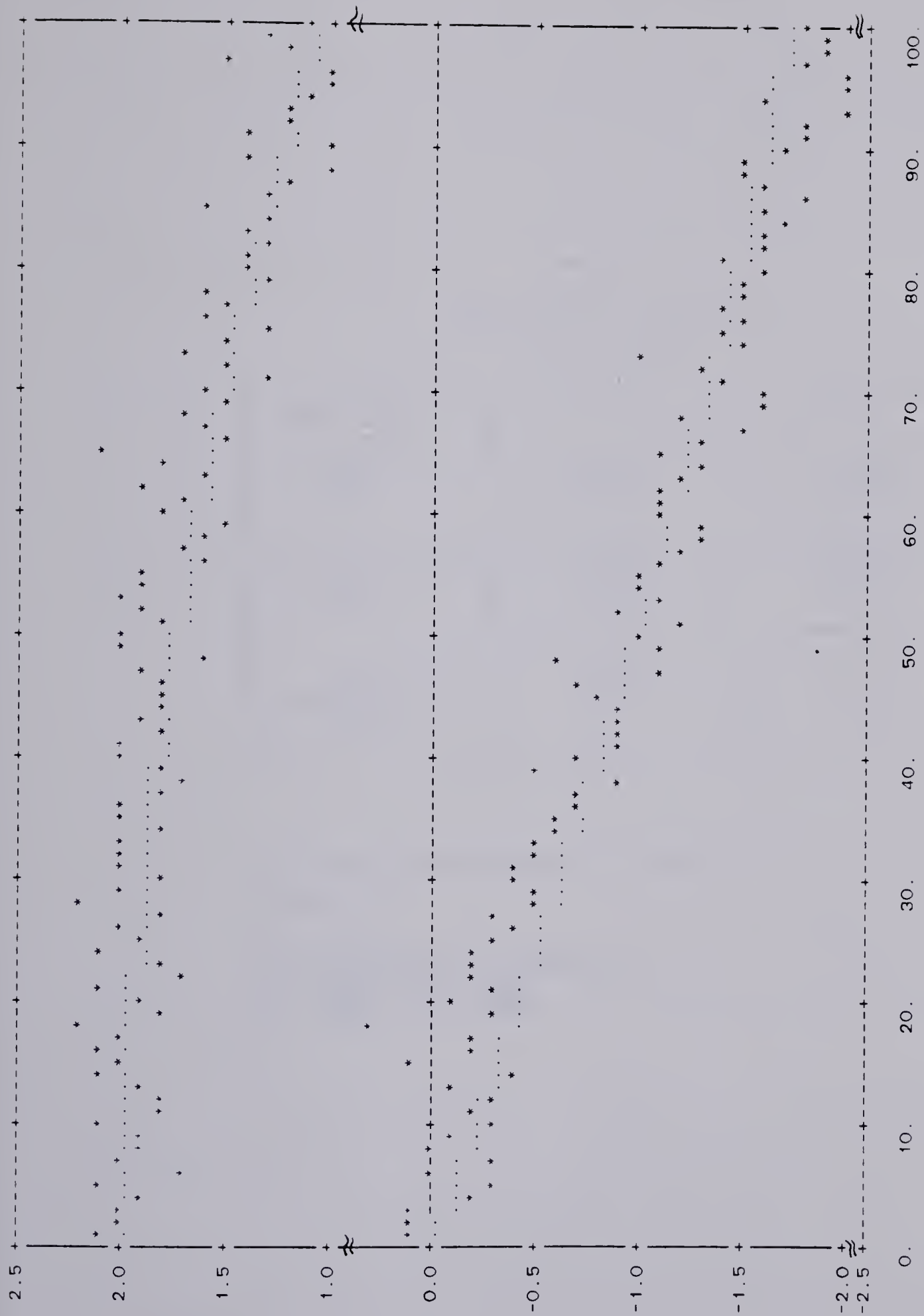


FIGURE X-7: Two-dimensional model: trajectory with and without noise inputs.

TABLE X-6
Two-dimensional Oscillator
OLS Estimates

Conditions	A1=0.01	A2=0.01
Q=1% R=10%	0.00367 (0.019) $R^2=0.0008$	0.0103 (0.0108) $R^2=0.0115$
Q=10% R=10%	-0.0898 (0.058) $R^2=0.0256$	-0.0054 (0.011) $R^2=0.0073$
Q=1% no measurement noise	0.0100 (0.00) $R^2=1.0$	-0.01 (0.0) $R^2=1$

The standard deviations are in brackets

Conditions:

100 data points

The seed numbers for GGNOF are:

U1 = 14785 U2 = 46583

V1 = 63976 V2 = 74372

The estimate of A2 is much better but it is also very insignificant.⁶¹

Running the program with 10% equation noise results in a drastic worsening of the estimates (A1 now has the wrong sign) as also shown in Table X-6. The estimates of A1 and A2 with only equation noise are shown for comparison.

3. LIKALM estimates

A first run was attempted using the LIKALM program to estimate the two unknown parameters A1 and A2. In this first run, the actual values of the parameters were used as a starting point for the iteration. The covariance matrices of the equation and measurement noises Q and R were assumed known, diagonal, with diagonal elements:

$$\text{for } Q: q_{11} = q_{22} = 0.000225$$

$$\text{for } R: r_{11} = r_{22} = 0.0225$$

The program also needs starting values for the Kalman Filter: the initial value of the state vector XE and its covariance matrix SE.

■ For this first run, XEo was taken as the actual starting value of the model and therefore since XEo is known with certainty, SEo = 0.

Under these perfectly ideal conditions, the program converges to:

$$A1 = 0.00794 \pm 0.0018$$

$$A2 = -0.00910 \pm 0.00017$$

⁶¹ These estimates are very bad, but Table X-12 presents estimates with different noise sequences that are better but still have the same standard deviations

with a value of the Negative Likelihood function of:

$$L = -298.4.$$

An approximation to the standard deviations is given by the square roots of the diagonal elements of the inverse of the Information matrix. Two iterations were necessary to obtain convergence. (Convergence is decided manually and was obtained when the Likelihood function showed no change and when the elements of the incremental step vector were smaller than a specified value chosen as 1% of each parameter value). The cost of running the program for this model with 100 data points is \$0.40 per iteration.

These values are much better than the OLS values indicated in Table X-5, and although obtained under very ideal conditions, demonstrate the capacity of the LIKALM estimation to improve on the traditional OLS estimation for this two-dimensional oscillator. The standard deviations also indicate acceptable parameter estimates.

Having obtained convergence under these ideal conditions, the performance of the method (robustness) should now be tested by relaxing them one by one in order to investigate their effect on the estimates. More specifically:

1. With ideal (actual) values for Q , R , X_{E0} and $SE0$, try different starting point for the parameter vector and draw a zone of convergence
2. With ideal Q and R , OLS starting values for A , assess the effect of using the first available data point for X_{E0}

and $SE0 = R$

3. Estimate Q and R along with A .
4. With XEO , $SE0$ and A as previously defined, asses the effect of wrong values for Q and R .

This last approach has been used here because, as explained in section (X-B.1), the estimation of Q and R has proven problematic.

Tables X-7 to X-11 present the results of these runs using the same noise sequence as for the OLS estimates. The effect of using different noise sequences is illustrated in Table X-12. The results reported in this last table are for the worse conditions that is:

$Ao = OLS$

$Q = R = \text{Sample covariance matrix, } S$

$XEO = Z1$

$SEo = R$

4. Analysis of results

■ Table X-7 shows that for this particular linear model, the program converges to the same final values for a very wide range of starting points in only 3 to 4 iterations (all other values being kept ideal). The worst case (which still allows for convergence) is when $A1$ and $A2$ take ten times their actual values, with the wrong sign! As can be seen from this table, there was no convergence for the last two values tried ; however, the conditions in these cases were rather drastic (a hundred times $A1$ and $A2$, with the wrong sign and $A1 = 5$, $A2 = 0$).

TABLE X-7
Two-dimensional Oscillator
LIKALM Estimates
Convergence zone under ideal conditions

A1	A2	A1	A2	Likelihood
0.01	-0.01	0.00794 (0.00187)	-0.00910 (0.00090)	-298.4
0.00367	-0.0103	0.00794 (0.00187)	-0.00910 (0.00090)	-298.4
0.05	0.0	0.00794 (0.00187)	-0.00910 (0.00090)	-298.4
0.05	0.05	0.00794 (0.00187)	-0.00910 (0.00090)	-298.4
-0.1	0.1	0.00794 (0.00187)	-0.00910 (0.00090)	-298.4
-1.0	1.0	divergence		
5.0	0.0	divergence		

Conditions:

All ideal
100 data points

The conclusion for this set of runs is that the algorithm is not very sensitive to the choice of a starting point. A bad OLS starting value is good enough to obtain convergence in this case. This is due to the fact that the model itself is very stable (linear oscillator) and not very sensitive to parameter variations as can be seen from Figure X-8 which illustrates the behavior of the model without noise inputs in the original model and if the parameters are replaced by their OLS estimates.

- Table X-8 demonstrates the effect of using the first data point instead of the unavailable initial state $\underline{x}_{E0}$, and to set $\underline{S}_{E0}$ (the covariance matrix of the random variable $\underline{x}_{E0}$) equal to R , the covariance matrix of the noise.

This choice makes sense since the measurements ($\underline{z}(1)$ in this case) consist of the true value of the state vector $\underline{x}(1)$, plus noise. The error covariance $\underline{S}_{E(1/1)}$ is then given directly as:

$$\begin{aligned} \underline{S}_{E(1/1)} &= E\{[\underline{x}(1) - \underline{x}(1/1)][\underline{x}(1) - \underline{x}(1/1)]'\} \\ &= \text{diag}(r_{11}, r_{22}) \end{aligned} \quad (34)$$

where r_{ii} is the i -th diagonal element of the measurement noise covariance matrix R . (see Mehra [84]).

The parameter values in Table X-8 along with their standard deviations show little change from the ideal case: the minimum attained is now a little higher ($L = -297.23$). In this run, the OLS starting values were used.

- Estimation of Q and R .

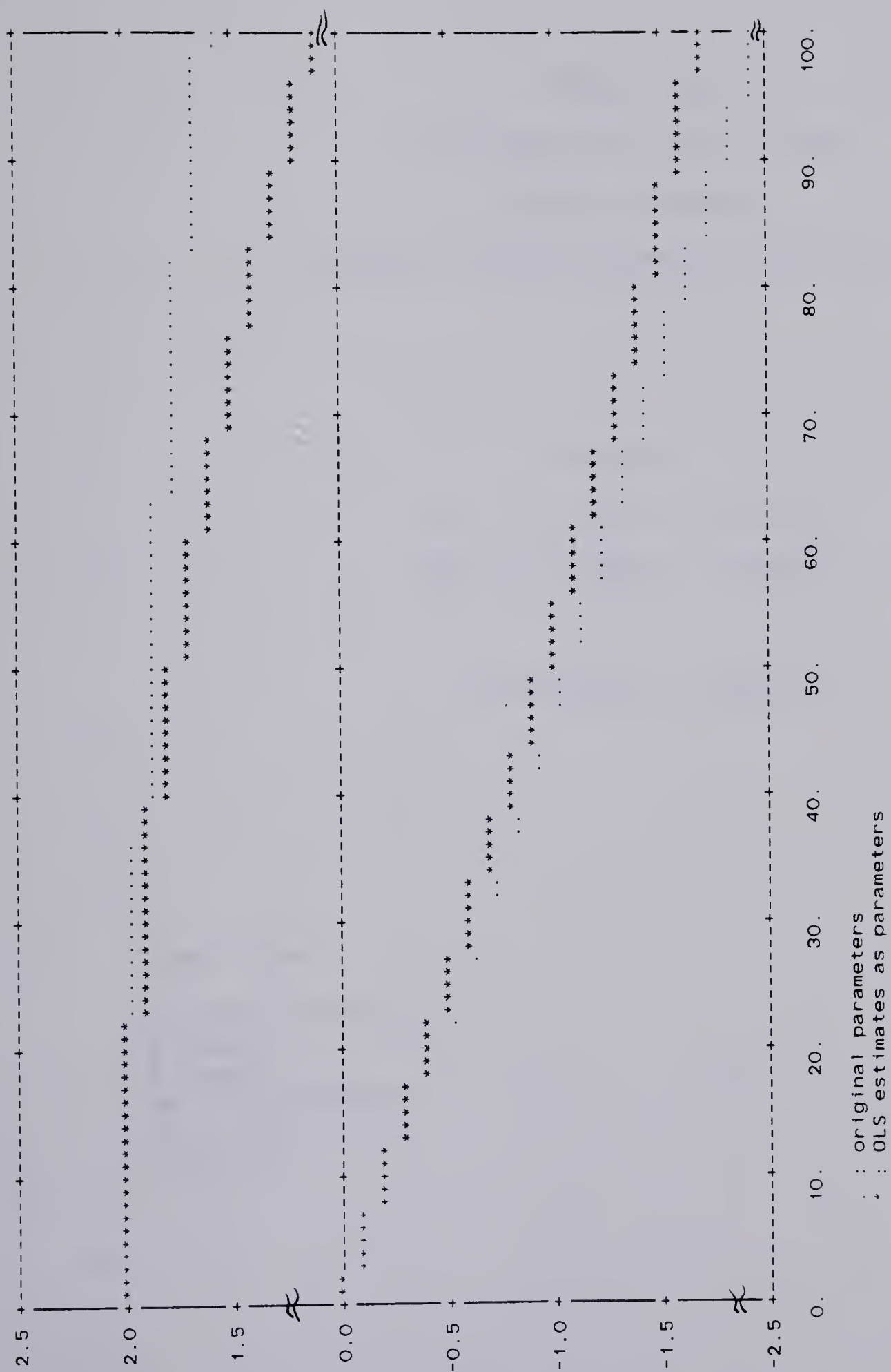


FIGURE X-8: Two dimensional model: sensitivity of model to changes in the parameters.

TABLE X-8
Two-Dimensional Oscillator
LIKALM Estimates
Wrong initial values for XEo and SEo

Results

$$A1 = 0.00787 \pm 0.00187$$

$$A2 = -0.00959 \pm 0.00096$$

$$\text{Likelihood} = -297.23$$

Conditions

100 data points

Q ideal

R ideal

Ao = OLS values

XEo = Z1

SEo = R

Having estimated the parameter vector A with Q and R fixed and specified at their ideal theoretical values, the estimation of Q and/or R along with that of A was attempted in order to confirm the one-dimensional results.

All attempts at estimating Q and/or R with A fixed or also estimated, using the first form of the information matrix, M_1 , led either to divergence problems or to a nonpositive definite information matrix, and the ensuing computational problems. This confirms the results obtained with the one-dimensional model, where the second derivative of the Likelihood function was negative at the minimum.

The use of the second form of the information matrix, M_2 , led to much better results. Since the noise levels were chosen equal for both equations, Q and R were taken as diagonal with equal diagonal elements:

$$q_{11} = q_{22} = 0.000225$$

$$r_{11} = r_{22} = 0.0225$$

The estimation results are reported in Table X-9.

The best minimum attained by the negative Likelihood function is when Q is allowed to be negative, which confirms our one-dimensional results. On the other hand, the estimation of R along with A yields a very good value for R , specially if we consider that the sample covariance matrix of the measurement noise (with hundred data points) is equal

TABLE X-9
Two-dimensional Oscillator
Estimation of A, Q and R

Estimated	Initial values	A1 = 0.01	A2 = -0.01	A3 = 0.00022	A4 = 0.0225	Like.	# iter.
A, Q	All ideal	0.00711 (0.00009)	-0.00847 (0.00005)	-0.00006 (0.0)	-	-329.7	4
A, R	All ideal	0.00786 (0.00183)	-0.00911 (0.00089)	-	0.01696 (0.00127)	-302.8	4
A, Q, R	All ideal	0.00822 (0.00119)	-0.00907 (0.00053)	0.00007 (0.00006)	0.01683 (0.00128)	-303.9	6
A, Q, R Q cons- trained	A0 = OLS Q0 ideal R0 ideal XE0 = Z1 SE0 = R0	0.00849 (0.00080)	-0.00950 (0.00033)	0.00001 0.00003	0.01741 (0.00129)	-302.9	5

Conditions:

100 data points
Q = 1%, R = 10%

to 0.016 (instead of the theoretical value of 0.025).

If Q is constrained to remain positive, the estimation of A , Q and R with ideal or approximate values for X_{E0} and S_{E0} , and a starting point at the OLS values, yields good estimates for A and Q . But Q is set at zero during the convergence process and ends up with a small and insignificant positive value.

- If Q and R present estimation problems, another approach might be to choose "ad-hoc" values known "a priori" as is the usual practice. However, these values might be wrong.

Table X-10 illustrates the effect of choosing wrong values for Q and R (taken as fractions or multiples of their actual values); in all these runs A_0 is set at its OLS value, X_{E0} and S_{E0} are ideal. The program still converges in 3 iterations in all the runs tried. However, the final values are now affected by the choice of Q and R .

As far as the algorithm is concerned, the equation noise seems to be irrelevant. The best estimate is obtained when R is equal to its true value and Q is smaller than the true value. However, variations in R when Q is set at its true value affect the Likelihood function drastically. The fact that the Likelihood function is much more sensitive to R than to Q is due to the fact that Q is not identifiable as was shown by the one-dimensional example. Thus this experiment again verifies our previous results.

- Table X-11 shows the effect of varying the ratio Q/R

TABLE X-10
Two-dimensional Oscillator
LIKALM Estimates
Wrong values for Q and R

Conditions	A1 = 0.01	A2 = -0.01	Likelihood
R' = R Q' = 10Q	0.00740 (0.00528)	-0.00899 (0.00276)	-285.9
R' = R Q' = R	0.00689 (0.01616)	-0.00889 (0.00878)	-242.2
R' = 10R Q' = Q	0.00874 (0.00476)	-0.00896 (0.00170)	-139.0
R' = 100R Q' = Q	0.00972 (0.00167)	-0.00884 (0.00087)	82.32
R' = 0.1R Q' = Q	0.00741 (0.00253)	-0.00899 (0.00101)	47.41
R' = 10R Q' = 10Q	0.00648 (0.0515)	-0.00874 (0.0275)	-49.12
R' = R Q' = 0.1Q	0.00874 (0.0008)	-0.00896 (0.00032)	-300.06

Conditions:

Ao = OLS
XEo, SEo: ideal

based on the approximation $Q + R = S$, as explained in section (X-B.1). Setting $R > Q$ improves the standard deviation of the estimates (which is a criterion more suitable for comparison than the actual parameter values [84]). Setting $R=100 \bullet Q$ (that is, the measurement noise equal ten times the equation noise, which is the actual experimental condition) drastically improves the standard deviations, with only slight improvement if Q is set smaller. This indicates to the modeler that the major error in this case is probably due to measurement noise. The modeler would probably choose:

$$A1 = 0.00856 \pm 0.00071$$

$$A2 = -0.00947 \pm 0.00035$$

which have the smallest standard deviations. These values are within two standard deviations of the true values and give the best minimum for the Likelihood function.

■ Finally, Table X-12 demonstrates the effect of using different noise sequences with $A_0 = \text{OLS values}$, $X_{E0} = Z(1)$, S_{E0} , Q and R set as indicated in the previous experiment using the values of S and $S/100$.

All OLS estimates are insignificant and all are improved by the LIKALM procedure. Although the values of the estimates varies, the standard deviations of the estimates are comparable for all the noise sequences.

5. Conclusion

It is apparent from this simple linear example that the LIKALM procedure drastically improves the estimates of

TABLE X-11

Two-dimensional Oscillator

LIKALM Estimates

Different Combinations of Q and R based on
the Least Squares Residual S

Conditions	A1 = 0.01	A2 = -0.01	Likelihood
Q = S = 0.034 R = S = 0.034	0.00686 (0.01987)	-0.00906 (0.01072)	-214.43
Q = 0.017 R = 0.017	0.00690 (0.01404)	-0.00907 (0.00758)	-256.44
Q = 0.03 R = 0.003	0.00518 (0.01843)	-0.00968 (0.01000)	-244.88
Q = 0.003 R = 0.03	0.00739 (0.00609)	-0.00941 (0.00325)	-271.72
Q = 0.0003 R = 0.03	0.00787 (0.00216)	-0.00958 (0.00111)	-286.08
Q = 0.00003 R = 0.03	0.00837 (0.00099)	-0.00951 (0.00049)	-289.13
Q = 3×10^{-6} R = 0.03	0.00854 (0.00074)	-0.00948 (0.00036)	-289.66
Q = 3×10^{-7} R = 0.03	0.00856 (0.00071)	-0.00947 (0.00035)	-289.70

Conditions:

100 data points

Ao = OLS

XEo = Z1

SEo = R

TABLE X-12

Two-dimensional Oscillator

OLS and LIKALM estimates with different noise sequences.

Conditions	A1		A2	
	OLS	LIKALM	OLS	LIKALM
1	0.00368 (0.01977) DW=2.97 R ² =0.0008	0.00787 (0.00216)	-0.01035 (0.01035) DW=3.1 R ² =0.0115	-0.00958 (0.00111)
2	0.01346 (0.02165) DW=3.12 R ² =0.0006	0.01157 (0.00244)	-0.01336 (0.01318) DW=2.99 R ² =0.0107	-0.01093 (0.00164)
3	0.01054 (0.02165) DW=2.89 R ² =0.0029	0.01151 (0.00222)	-0.00835 (0.013108) DW=3.02 R ² =0.0000	-0.01037 (0.00150)
4	0.00933 (0.01803) DW=3.01 R ² =0.0006	0.00953 (0.00215)	-0.01140 (0.012060) DW=3.23 R ² =0.0009	-0.01131 (0.00128)

Conditions:
 100 data points
 A0=OLS
 XE0=Z(1), SE0=R=S
 Q=S/100

parameters and their confidence intervals, in the presence of measurement noise, if the modeler is in a position to specify reasonable bounds for the covariance matrices. There is an identification problem if we attempt to estimate all the unknown parameters (including the covariance matrices) since the data we have is insufficient to provide for the estimation of all the parameters.

Having gained some confidence in the performance of the program for this simple case, it was then attempted to apply the method to a nonlinear case.

B.3 Three-Dimensional Nonlinear Model

The cost of running the Market-Growth model being very high, the following three dimensional model was chosen to determine the behavior of the algorithm under a specification very similar to the Market-Growth model.

1. The model

The equations of this three dimensional nonlinear model are, in their FORTRAN form:

$$TX1 = X1 + DT * (A1*X1*X2 + A2*X2^2/X3 + U1) \quad (35)$$

$$TX2 = X2 + DT * (A3*X3/X1 + A4*X2 + U2) \quad (36)$$

$$TX3 = X3 + DT * (A5*X1/X2^2 + A6*X2 + U3) \quad (37)$$

and:

$$Y1 = TX1 + V1 \quad (38)$$

$$Y2 = TX2 + V2 \quad (39)$$

$$Y3 = TX3 + V3 \quad (40)$$

where:

$$A1 = -0.001326$$

$$A2 = 0.0424$$

$$A3 = 34.57$$

$$A4 = -1.45$$

$$A5 = 4.3217$$

$$A6 = 0.4187$$

with U_i and V_i random sequences generated by GGNOF with different seed numbers and standard deviations equal to 1% and 10% of the mean of the relevant level variables respectively.

The starting values of the model are:

$$X1_0 = 4.$$

$$X2_0 = 150.$$

$$X3_0 = 1.$$

The model exhibits a nonlinear behavior illustrated by Figure X-9. with $DT = 0.2$

Since the equation noise is also multiplied by DT in the data generating model, the standard deviations of the equation noise sequences in the GGNOF functions were adjusted by dividing them by DT . The theoretical (rounded off) values for Q and R are:

$$q_{11} = 2.25$$

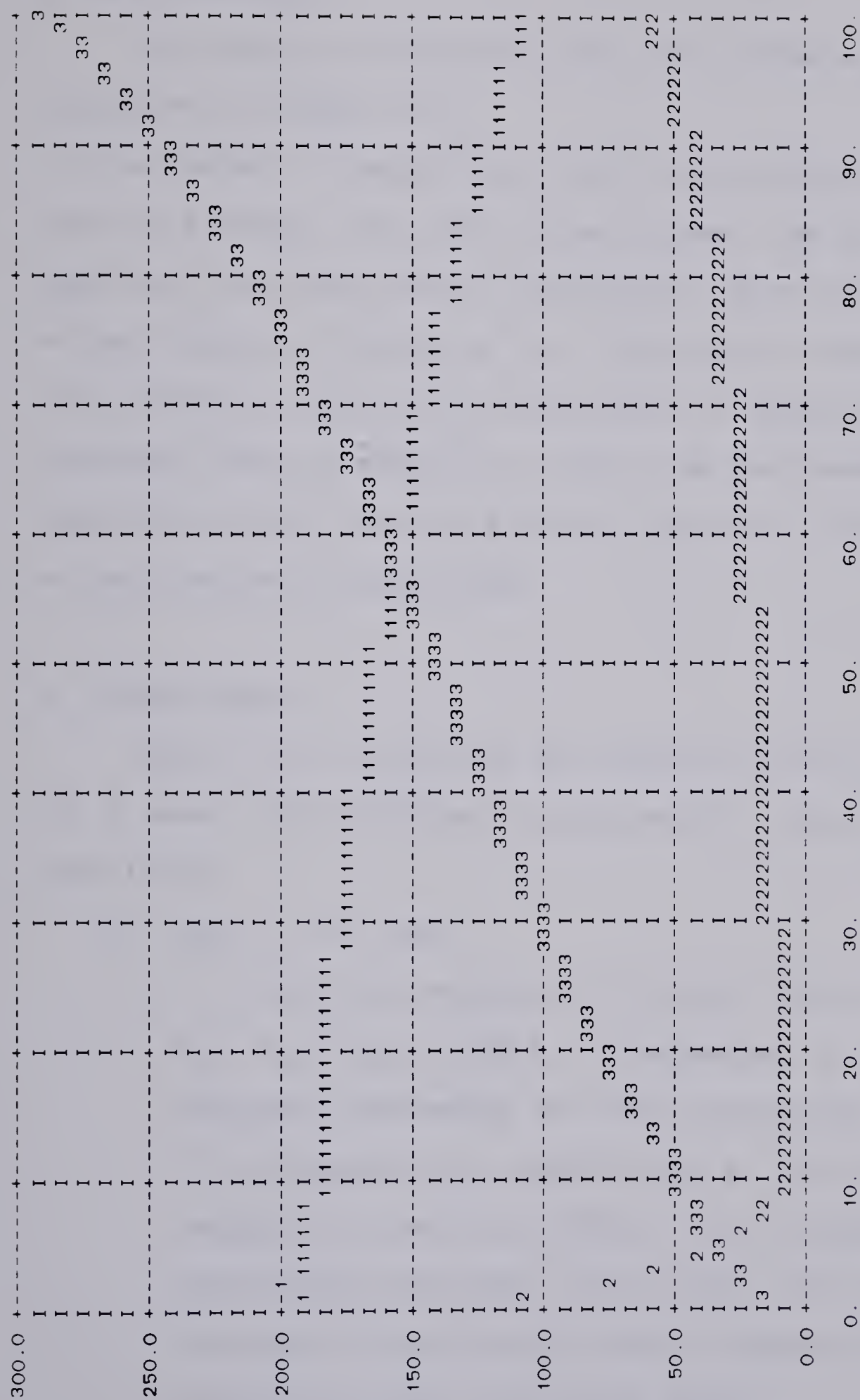
$$q_{22} = 0.0625$$

$$q_{33} = 2.25$$

and

$$r_{11} = 225.$$

$$r_{22} = 6.25$$



$$r_{33} = 225.$$

2. OLS estimates

The results of the OLS runs with 100 data points are presented in Table X-13.

All parameters in equations 1 and 3 are insignificant and deviate strongly from their true values. The parameters in equation 2 are much better and deviate from their true values only by a factor of 1.2. Therefore, the modeler would have probably accepted the estimates in equation 2 and rejected those in equation 1 and 3. He has however no way of improving on the specification of equation 1 and 3 that are already perfectly specified.

3. LIKALM runs

Table X-14 illustrates the results of the LIKALM runs for 4 cases corresponding to progressive relaxation of ideal conditions.

a. Run 1: all ideal

When the program is run with ideal values for A_o , XE_o , SE_o , Q and R , it converges to very good parameter estimates in three iterations.

All parameters are significant at the 95% level of probability and fall within 1 to 2 standard deviations from their true values. The program therefore theoretically has the capacity of estimating such a nonlinear model.

b. Run 2: OLS starting values

TABLE X-13

Three-dimensional model

OLS estimates

Parameter	True value	OLS (st. dev.)	R ² DW SSR
A1	-0.001326	-0.0030 (0.0026)	0.0193 3.04 59469.
A2	0.0424	0.1735 (0.1685)	59469.
A3	34.57	29.07 (2.42)	0.6032 2.68
A4	-1.45	-1.27 (0.11)	1144.
A5	4.3217	1.1262 (10.906)	0.0010 2.92
A6	0.4187	0.4190 (0.2939)	47813.6

Conditions:
100 data points
 $Q=1\%$, $R=10\%$

```
Noise sequence seeds: U1 = 43567      V1 = 36597
                      U2 = 56493      V2 = 21485
                      U3 = 52486      V3 = 724195
```


TABLE X-14
3-dimensional model
LIKALM estimates

Conditions	A1	A2	A3	A4	A5	A6	Neg. Likelihood
Real values	-0.001326	0.0424	34.57	-1.45	4.3217	0.4187	
OLS values	-0.00296 (0.00257)	0.1735 (0.1685)	29.07 (2.42)	-1.27 (0.11)	1.1262 (10.9006)	0.4190 (0.2939)	
1 All ideal	-0.00170 (0.00002)	0.04245 (0.00125)	34.65 (0.92)	-1.48 (0.03)	4.8459 (1.0256)	0.46615 (0.0267)	796.8
2 Ao=OLS All other ideal	-0.00170 (0.00002)	0.04245 (0.00125)	34.65 (0.92)	-1.48 (0.03)	4.8459 (1.0256)	0.46615 (0.0267)	796.8
3 Ao=OLS XEO=2.0 SEO=R Q,R ideal	-0.00175 (0.00020)	0.06925 (0.04590)	27.65 (0.92)	-1.17 (0.03)	5.7084 (1.1099)	0.4633 (0.0289)	834.9
4 Ao=OLS XEO=Z1 SEO=R Q=S/100 R=S	-0.0018- (0.0004)	0.07070 (0.0802)	27.65 (1.33)	-1.17 (0.05)	5.9338 (1.7569)	0.4593 (0.0467)	860.0

Conditions:
Q=1%, R=10%
100 data points

The algorithm converges to the same final values as in run 1.

c. Run 3: $A_0 = \text{OLS}$, $X_{E0} = Z$ and $S_{E0} = R$.

The final values are now worse (specially for equation 2 where OLS values are comparable) but, except for equation 2, all fall within 1 to 2 standard deviations of their true values.

d. Run 4: $A_0 = \text{OLS}$, $X_{E0} = Z$, $S_{E0} = R$, $R = S^{62}$ and $Q = S/100$.

The final values are very similar to those obtained for run 3, but the standard deviations are now much larger expressing a greater uncertainty in the quality of the estimates. This is to be expected since the values of S are:

$$S_{11} = 594.$$

$$S_{22} = 11.4$$

$$S_{33} = 478.$$

so that even with a perturbation of about 100% in the values of Q and R , the algorithm still converges.

4. Conclusion

From the above results with the nonlinear three-dimensional model, it appears that the LIKALM program has the capacity to improve drastically over OLS estimates of a nonlinear model in the presence of equation and measurement noise. Although the behavior of

⁶²See page 169

this three-dimensional model is not as unstable as that of the Market-Growth model, the nonlinearities present here are of the same kind and can be taken as an example of the nonlinearities generally found in system dynamics models.

The instability of the model, the singularities that appear are typical of what can be expected from a system dynamics model.

Although Q and R were not estimated along with the other parameters, the use of the residual noise process covariance matrix S as an approximation for $R + Q$, shows that a wide range of variation in their values is still possible and yields acceptable results.

A whole range of new problems is to be expected as the size of the model approaches those usually encountered in "Real-Life" models. The following experiments with the Market-Growth model illustrate this fact very well (although this model is still of modest size)

B.4 The Market-Growth model

As it is written, the Market-Growth model does not directly lend itself to estimation in the state-space form since equation (VII-8) is not a state equation but merely an identity⁶³. The variable DDC had to be substituted into equation (VII-4) (the only equation where DDC appeared) thus reducing the system to nine equations. The parameter values

⁶³ See equations of the model page 77

in equation (VII-4) are therefore changed and the new form of the equation is as follows:

$$\begin{aligned} dPC001 = & DT \bullet [K' 14 \bullet PC(-DT) + K' 15 \bullet PC(-DT) \bullet DDRC(-DT) \\ & + K' 16 \bullet PC(-DT) \bullet DDRC(-DT)^2 + K' 17 \bullet PC(-DT) \bullet DDRC(-DT)^3 \\ & - 0.25 \bullet PC001(-DT) + N4] \end{aligned} \quad (41)$$

Where:

$$K' 14 = -0.11517$$

$$K' 15 = 0.09026$$

$$K' 16 = -0.02643$$

$$K' 17 = 0.00338$$

All the other equations remain the same.

1. OLS runs

Although the model has been thoroughly estimated with OLS in section (VII-B) new OLS runs had to be tried with the new form of equation (4) and the new data generated by the model. The data generated by the two versions of the model differ slightly due to different noise sequences.

The noise sequences were generated as indicated in Table X-15.

The OLS results with this new version of the Market model are presented in Table X-16 and can be compared to those indicated in Tables VII-8 to VII-11.

Since the two data sets differ, the estimates for all equations are not the same as previously. Estimates for K1, K2 and K3 are now significant and better than previously estimated. All estimates for equation 2 are

TABLE X-15

Noise Sequences for Market model

Seed numbers for GGNOF routine

Equation noise	Measurement noise	for
947312	67751	BL
2345	93023	DRA
6312	10735	S
95642	76925	PC001
44872	5675	DDRC
443521	22768	DDRM
	31356	PC

TABLE X-16

Market model

OLS estimates with new form of equation 4.

Parameter	True value	Old value	New value	st.dev.	t	R ² (R ²) DW
K1	475	221.2	345.3	91.9	3.76	
K2	-61.5	-26.8	-46.65	12.0	-3.89	0.1591
K3	-0.6178	-0.4151	-0.4455	0.18	-2.42	0.9619
K4	0.1324	0.1123	0.0754	0.04	1.85	2.77
K5	-0.00975	-0.0100	-0.0044	0.0033	-1.33	
K6	0.6178	0.4513	0.3317	0.0599	5.54	0.2440
K7	-0.1324	-0.1017	-0.0086	0.0136	-5.03	0.9260
K8	0.00975	0.00782	0.00485	0.00112	4.32	2.64
K9	-1.0	-0.7032	-0.5331	0.0981	-5.44	
K10	3.10×10^{-4}	11.74×10^{-4}	11.74×10^{-4}	2.8×10^{-4}	4.19	0.1403 0.9652
K11	-0.05	-0.255	-0.255	0.065	-3.93	2.74
K12	-0.11517		0.0782	0.0547	1.43	0.8459
K13	0.09026		-0.0884	0.0486	-1.82	0.9905
K14	-0.02643		0.0263	0.01388	1.90	2.27
K15	0.00338		-0.00159	0.00126	-1.26	

Conditions:

100 data points

Q=1%, R=10%

t is the t-statistic

(R²) is for estimation in the first difference form (see section VII)

still significant but worse than their previous counterparts. The values of K10 and K11 in equation 3 are identical in both cases. The coefficients in equation 4 cannot be compared since they now have different values. However, they still all have the wrong sign and are all insignificant. Overall the OLS estimates obtained for this single run confirm the results obtained earlier as a verification of Senge's results (see section (VI-B.2)).

2. LIKALM runs

In order to avoid computational problems with the LIKALM program (due to the great number of matrix operations and the wide disparity in variable values⁶⁴), the data was scaled as follows:

$$BL' = BL/100$$

$$DRA' = DRA/100$$

$$S' = S$$

$$PC001' = PC001/10$$

$$PC002' = PC002/10$$

$$PC003' = PC003/10$$

$$PC' = PC/100$$

$$DDRC' = 10*DDRC$$

$$DDRM' = 10*DDRM$$

The new parameter values to be estimated are presented in Table X-17.

OLS estimates with this new scaled data are very

⁶⁴ Ranging from 0.4 to 120000

TABLE X-17
 Market-Growth Model
 Scaled parameter values

Parameter	True value	Scaled value
K1	475	4.75
K2	-61.5	-0.00615
K3	-0.6178	-0.6178
K4	0.1324	0.1324
K5	-0.00975	-0.00975
K6	0.6178	0.6178
K7	-0.1324	-0.1324
K8	0.00975	0.00975
K9	-1.0	-1.0
K10	3.10×10^{-1}	0.03
K11	-0.05	-0.05
K12	-0.11517	-1.1517
K13	0.09026	0.09026
K14	-0.02643	-0.002643
K15	0.00338	0.000034

similar (but scaled) to those presented in Table X-16 and can be found in Table X-18.

The cost of running the LIKALM program with the entire Market-Growth model being very high⁶⁵, only 50 data points were used in the following experiments.

As a first experiment, the equations were estimated one at a time, under ideal conditons⁶⁶ all other parameters being fixed at their true values.

The estimation results are presented in Tables X-19 to X-21.

In all cases, the program converges to values that are very close to and, in most cases, within one standard deviation of the true values. The accuracy of this method as compared to OLS is striking, in particular for equations 3 and 4. Equation 4, however seems to present some convergence difficulties, and although the estimates are much better than in OLS, K14 and K15 are still unacceptable even though the standard deviations seem to indicate little error.

Having thus established the capacity of the program to converge to a suitable minimum with this complex and highly nonlinear model, for each individual equation, estimation of the parameters in the first three equations was then attempted.⁶⁷

⁶⁵ \$60 per iteration for 50 data points

⁶⁶ See section (X-B.2) for definition of "ideal"

⁶⁷ Equation 4, which had dispalyed a bad behavior in the first runs, was left out here

TABLE X-18

Market model

OLS estimates with new form of equation 4.
Scaled data

Parameter	True value	OLS	t	R ² (R ²) DW
K1	4.75	3.73	3.98	
K2	-0.0615	-0.0506	-4.1	0.1729
K3	-0.6178	-0.4061	-2.28	.9625
K4	0.1324	0.0645	1.64	2.74
K5	-0.00975	-0.00367	-1.13	
K6	0.6178	0.3317	5.53	.2440
K7	-0.1324	-0.0686	-5.03	.9260
K8	0.00975	0.00485	4.32	2.64
K9	-1.0	-0.53	-5.45	
K10	0.03	0.007	4.19	.1403 .9652
K11	-0.05	-.255	-3.93	2.74
K12	-1.1517	0.7818	1.43	.8459
K13	0.09026	-0.08838	-1.82	.9905
K14	-0.002643	0.002631	1.90	2.27
K15	0.000034	-0.000016	-1.26	

Conditions:

100 data points

Q=1%, R=10%

t stands for t-statistic

TABLE X-19

Market model

LIKALM estimates of parameters in equation 1 only.

Parameter	True value	LIKALM	st.dev.
A1	4.75	4.767	0.008
A2	-0.0615	-0.0635	0.0008
A3	0.6178	-0.6331	0.0165
A4	0.1324	0.1320	0.0004
A5	-0.00975	-0.00941	0.00043

Conditions:

50 data points

, $Q=1\%$, $R=10\%$

All ideal

3 iterations

TABLE X-20
Market model
LIKALM estimates of parameters in equation 3 only,
under different conditions

Parameter	A10	A11
True value	0.03	-0.05
LIKALM (st. dev.)	0.042 (0.011)	-0.075 (0.026)
LIKALM 100 pts. (st. dev.)	0.0409 (0.0066)	-0.074 (0.015)
LIKALM 50 pts. Ao=OLS (st. dev.)	0.043 (0.012)	-0.075 (0.026)

Conditions:
Q=1%, R=10%
All ideal unless otherwise specified

TABLE X-21

Market model

LIKALM estimates of parameters in equation 4 only.

Parameter	True value	LIKALM	st.dev.
A12	-1.1517	-1.1535	0.00002
A13	0.09026	0.09029	0.00001
A14	-0.002643	-0.00404	0.00011
A15	0.000034	0.00010	0.00000

Conditions:

Q=1%, R=10%

50 data points

All ideal

The results of this run (still under ideal conditions), in Table X-22, do not indicate much change from the previous single equation runs.

The introduction of equation 4 in the set of parameters to be estimated does however tend to alter the estimation results and the rate and conditions of convergence.

Because of the high cost of the runs, only a few experiments were attempted with the entire model.

1. Runs with $Q=1\%$ and $R=10\%$, and ideally specified.
 - a. If the initial values are chosen as the OLS values from Table X-18, with $XE_0 = Z1$ and $SE_0 = R$, the estimation problem is so ill-conditioned that the algorithm runs consistently into divergence or numerical singularity problems.
 - b. If, all other conditions remaining the same, the signs of the parameters in equation 4 are changed to correspond to their true values, the algorithm still faces the same problems.
 - c. If now, the starting point is changed to the true values altered by 10% ($A - 10\%$), the algorithm converges after a slight overshoot, and the values after 5 iterations are reported in Table X-23 along with the true and OLS values. The successive values taken by the negative Likelihood function are:

TABLE X-22

Market model

LIKALM estimates of parameters in equations 1 to 3.

Parameter	True value	OLS	LIKALM	st. dev.
A1	4.75	3.73	4.76	0.0039
A2	-0.0615	-0.0506	-0.0645	2.00056
A3	-0.6178	-0.4061	-0.6350	0.0172
A4	0.1324	0.0645	0.1321	0.0003
A5	-0.00975	-0.00367	-0.00128	0.00043
A6	0.6178	0.3317	0.6061	0.00344
A7	-0.1324	-0.0686	-0.1328	0.0572
A8	0.00975	0.00485	0.01006	0.00089
A9	-1.0	-0.53	-0.99	0.2726
A10	0.03	0.117	0.046	0.01235
A11	-0.05	-0.255	-0.077	0.02627

Conditions:
 Q=1%, R=10%
 50 data points
 All ideal

TABLE X-23

Market model

LIKALM estimation of entire model: $A_0 = A - 10\%$

Parameter	True value	OLS (st. dev.)	LIKALM (st.dev.)
K1	4.75	3.73 (0.93)	4.30032 (0.00030)
K2	-0.0615	-0.0506 (0.0123)	-0.05506 (0.00003)
K3	-0.6178	-0.4061 (0.1783)	-0.54395 (0.00155)
K4	0.1324	0.0645 (0.0394)	0.10007 (0.00002)
K5	-0.00975	-0.00367 (0.00325)	-0.00759 (0.00006)
K6	0.6178	0.3367 (0.0599)	0.54925 (0.00007)
K7	-0.1324	-0.0686 (0.0136)	-0.11310 (0.00035)
K8	0.00975	0.00485 (0.00112)	0.00827 (0.00002)
K9	-1.0	-0.53 (0.10)	-0.90009 (0.00001)
K10	0.03	0.007 (0.028)	0.02803 (0.00017)
K11	-0.05	-0.255 (0.065)	-0.04506 (0.00001)
K12	-1.1517	0.7818 (0.5471)	-1.00308 (0.00125)
K13	0.09026	-0.08838 (0.04865)	0.09112 (0.00066)
K14	-0.002643	0.002631 (0.001388)	-0.00324 (0.00007)
K15	0.000034	-0.000016 (0.000013)	0.00004 (0.00000)

Conditions:

Q=1%, R=10%

LIKALM: 50 data points, OLS: 100 data points

All ideal except A_0

2655, 1380, 1475, 1069, 1030,

Convergence is slow because, even though the gradient values are in the hundreds or thousands, the increments at each iteration are very small due to the small size of the elements in the inverse of the information matrix (of the order of 10^{-8} in the average). Thus the initial values are not much affected even after 5 iterations. The standard deviations are also correspondingly small.

2. Runs with $Q=1\%$ and $R=10\%$ but R specified at 10 times its value.

These runs were attempted to investigate the effect of misspecification of R and because previous experiments with the two-dimensional model showed that overspecifying R did affect the information matrix and consequently, the rate of convergence, much more than that of Q .

- a. With A_0 equal to the OLS values, all other initial values as in the previous experiments, the value of the negative Likelihood function on the first iteration is now much better than in the first run (1245 as compared to the previous 4723), thus indicating a better fit, but the algorithm then starts diverging and ends up with numerical singularity.
- b. The same happens if we start with $A = 10\%$.

A few other experiments with this model and with the smaller order models were tried to attempt to correct some of the numerical problems, but without much success. The number of experiments attempted was very much limited by the cost of the runs, and of course given the size of the information matrix and the results of the experiments with smaller models, the estimation of Q or R was not attempted.

3. As a matter of interest, a run was made with $Q=R=1\%$ but R specified as 10 times its value.

Starting with $A = 10\%$, the algorithm converges after 5 iterations to values very similar to the Least Squares values for this case. As expected, when the measurement noise is not high, the OLS and LIKALM values are comparable...

3. Conclusion

From these few experiments it appears that the convergence zone is much narrower than in the previous cases. The Likelihood surface is probably very sharp around the minimum. The instability of the model is a major limiting factor. It is actually a good proof of the robustness of the algorithm that it would still converge with $A_0 = A = 10\%$, given the behavior of the model with these parameter values, as illustrated by Figure X-10.

Due to the highly nonlinear character of the model, to its instability and high sensitivity to parameter

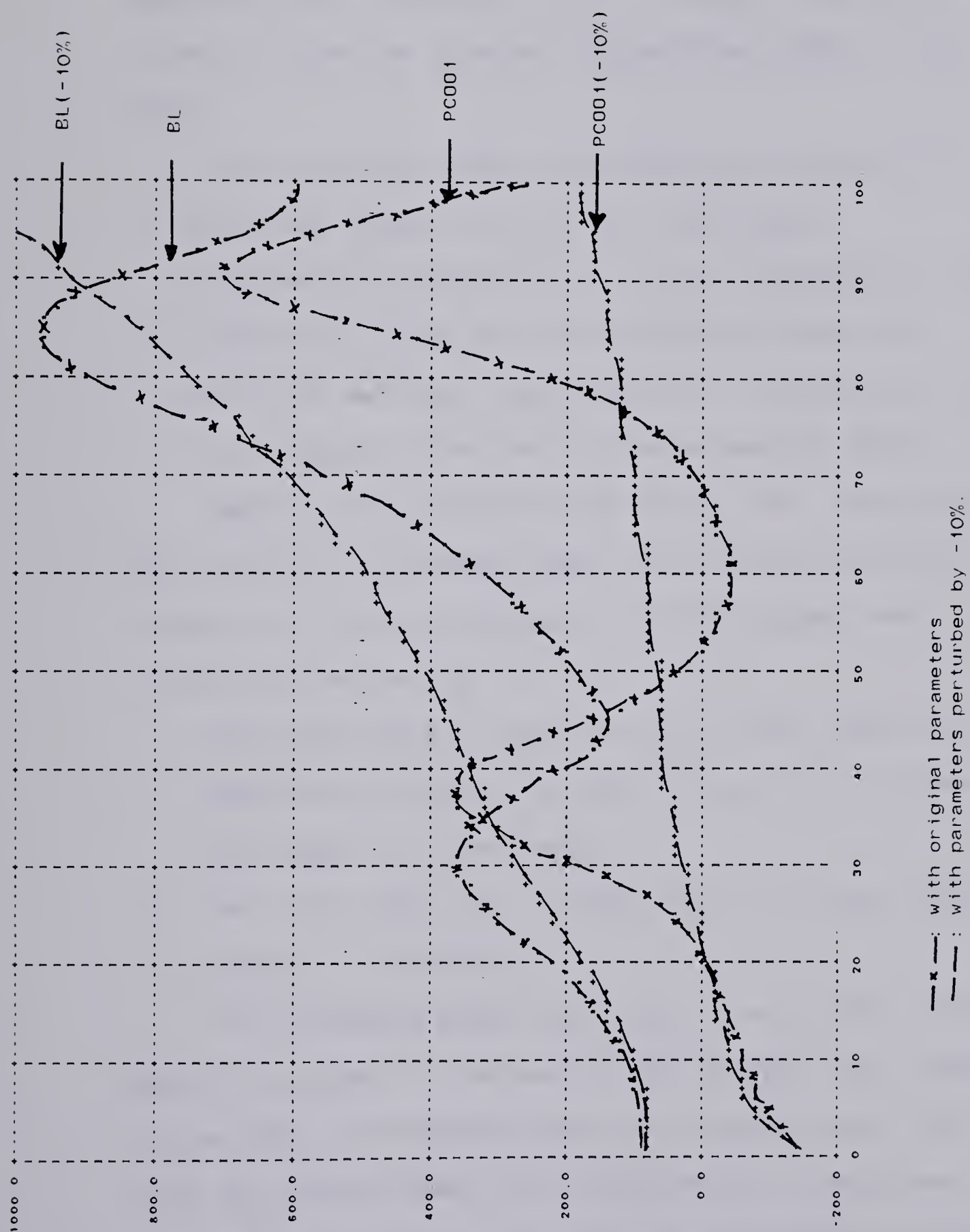


FIGURE X-10: Sensitivity of two selected variables of the Market-Growth model to a 10% change in all parameters.

changes, convergence of the program is dependent on the choice of starting values and encounters many computational problems. This is however a characteristic of many iterative gradient algorithms. Mehra [112] notes:

"The use of Maximum Likelihood estimation in practice leads to difficult nonlinear programming problems. It is not uncommon to find situation where the likelihood surface has multiple maxima, saddle-points, discontinuities, and singular Hessian in the parameter space..."

Mehra also points out in [112] that the method used here runs into problems when the information matrix is singular or nearly singular, in which cases two situations may arise:

1. The computed M^{-1} may turn out to be indefinite or negative definite, so that $J(A_{i+1}) > J(A_i)$ where J is the negative likelihood.
2. The step size, $\Delta A_i = \alpha_i \cdot M_i^{-1} \cdot g_i$ ⁶⁸ is very large in singular directions.

This happens when the first form of the information matrix is used⁶⁹. The use of the second form, however forces the information matrix to remain positive. This does not avoid numerical singularity in any case.

A few techniques have been suggested for handling singular or nearly singular information matrices and are

⁶⁸ See p. 135

⁶⁹ See section (IX-F.1)

presented in Mehra [112]. They all involve a great number of further computations, reductions in the order of matrices, etc..., with no guarantee of convergence. These are all complex problems of numerical analysis and given the cost of the runs and the number of runs that should be tried in order to obtain any satisfactory results, no further experiments were attempted.

This leaves us with quite a dissappointing picture of the performance of the LIKALM program in the case of the Market-Growth model.

XI. CONCLUSION

A. Introduction

The original purpose of this thesis was to attempt to bridge the gap between econometric and system dynamics model-building by providing a technique allowing for formal estimation of parameters in a nonlinear System Dynamics feedback model, in the presence of measurement noise. If such estimation proved satisfactory and if test statistics associated to it proved useful in determining acceptable confidence in parameter estimates and guided the modeler towards possible better specification, then the argument between econometric and system dynamics modelers (and in particular the issues of "arbitrary" parameters, lack of reliance on data in system dynamics and the requirements of model validation) would have less reason to exist.

No available econometric estimation technique proved capable of meeting the requirements of the estimation of parameters in a nonlinear dynamic feedback model in the presence of measurement noise since the emphasis of econometric research has been on simultaneous equations systems and models with perfect measurement. The recent shift to study the effect of measurement noise remains within the framework of linear simultaneous equations systems.

On the other hand, on the control side, the rapidly growing number and range of applications of the Filtering

Form of the Maximum Likelihood algorithm, which proved its capacity to perform very well in the context of noisy dynamic nonlinear models of physical systems, suggested that this approach might be useful in the field of socio-economic modeling and provided a base for its testing in this context.

Indeed this powerful technique has, in the various experiments conducted here, proven its capacity to drastically improve upon the Least Squares estimates, provided the models are small, perfectly specified, and sufficient "a priori" knowledge about the equation and measurement noise statistics is available.

Such detailed knowledge of both the process at hand and the noise characteristics exists in most engineering applications and this explains the success of the LIKALM approach in those fields.

If however, we now move to socio-economic models where the problems at hand are typically ill-suited for formal mathematical treatment, the number of variables and the parameters to be estimated several orders of magnitude larger than the typical physical applications, the degrees of freedom and uncertainties numerous, the measurement noise difficult to estimate and the residual processes in the model specification (which "explain" the uncertainties in functional relationships) large, one can no more assume "a priori" knowledge of the noise statistics nor near perfect specification.

The noise covariance matrices must then also be estimated and this entails various identification and numerical problems. Many limitations of the LIKALM algorithm become increasingly critical and preclude its use in socio-economic applications, and in particular in System Dynamics models.

Some of these limitations are outlined below.

B. Limitations of the LIKALM algorithm

1. One of the most important problems in the applicability of the Maximum Likelihood algorithm to large-scale systems is the great number of computational problems that occur. Mehra [112] deals at length with some of these aspects. Convergence and identifiability problems are typical of this kind of complex programs. Typical also are the great number of "ad-hoc" adjustments and the "hammering" that is needed to get the algorithm to converge, which depend very much on the talent of the experimenter and his detailed knowledge of both the model at hand and the program. A great fault of many complex "packages" of this kind is that, in order to allow for maximum generality and flexibility, thus containing several different search algorithms, for example, they become so intricate as to preclude their use by anyone other than their creator
2. Once the algorithm has converged, there is no guarantee that it has converged to the absolute maximum of the Likelihood function which is the only one for which the

estimate is proven to be unbiased, consistent and efficient.

3. All the theoretical results are for linear systems. The degree to which the "extended" Kalman Filter approximation is valid is not clear and needs to be investigated.
4. The use of the Kalman Filter is dependent upon the perfect specification of the model. What the values of the parameters mean in the case of imperfect specification is still to be examined.

Another limitation inherent to the Kalman Filter is that this formulation requires that all the state equations be included in the model of the system to be estimated. In other words, the whole system has to be entered in the program even if only a few parameters are to be estimated. A System dynamics model typically includes up to 200 state equations. All of these would have to be included in the formulation of the Kalman filter unless some way of separating strongly decoupled "subsystems" in the model can be devised.

5. Apart from the fact that the diagonal elements of the inverse of the information matrix provide estimates of the variance of the estimates, nothing much has been said here about various statistical tests of significance or overall fit and their applicability.
6. Last but not least, one of the limiting factors that assumed the greatest importance in the last experiment

is the cost of running the Market-Growth model.

As it stands, the program with the built-in "hard" limits of 10 equations and 40 parameters, because of the many matrix computations and the size of the matrices that appear in the sensitivity equations, requires over 1.5 million bytes of storage space. In the process of computing the information matrix, the four-dimensional matrix

$$SS = Sz^{-1} \bullet (dSz/dA) \bullet Sz^{-1} \bullet (dSz/dA)$$

of maximum dimensions (10,10,40,40) occupies by itself 640,000 bytes.

One iteration through 50 data points with the Market-model takes 114 CPU seconds which, at a low rate of \$15 for 20 CPU seconds at the University of Alberta, costs about \$60. A minimum of 3 to 4 iterations is needed to observe convergence, which puts a run to an average of about \$200.

This is using only 50 data points. The Maximum Likelihood method needs long data sequences, typically from 300 to 1000 data points, since all its properties are asymptotic.

This program was written without regard to any form of optimization. It can confidently be stated that the storage required, and consequently the cost of running the program, can be cut by a factor of 5 to 10.

It remains nonetheless that, even though this LIKALM program is very inefficient, the high cost is a general

feature of this approach. Mehra [84] reports a cost of 1 minute of CPU time per iteration for a model with four state-variables, 23 parameters and about 200 data points. Similar cost limitations are emphasized for the GPSIE program [85].

Even if all the above limitations were solved, the costs involved for a thorough formal estimation of a large-scale model would make the use of this algorithm extremely impractical, bearing in mind that one run of a 200 equation System Dynamics model with the Dynamo compiler costs only about \$2.

From the above limitations and the experience gained from the experiments, it appears that such formal parameter estimation, which to be consistent would have to allow for all the non-ideal situations described in Figure VI-2, is not the natural approach for the validation of System Dynamic models.

It might be useful at this point to restate briefly some significant aspects of the System Dynamics methodology.

The models are built with emphasis on real-world meaning for all parameters, on understanding model behavior in the light of the complex structure of feedback mechanisms rather than on prediction, intuitive understanding of the processes at work rather than on heavy reliance on data... These models still have to be validated through contact with data. However, in light of these characteristics, it certainly is more worthwhile investigating other means of

model validation, more suited to the modeling approach and the nature of the models at hand than to try to force a ill-suited approach into behaving properly.

C. Suggestions for Further Research

In their effort to demonstrate the fallacy of using single equation statistical tests for the selection of model variables, Mass and Senge [70]⁷⁰ show that such statistical tests "should not be viewed as tests of model specification *per se*, but as tests of a particular type of data usefulness: they warn the modeler when available data do not permit accurate estimation of a model parameter".

However, they illustrate the fact that a model relationship may be difficult to estimate yet extremely important for overall behavior and thus state that "tests which analyze the impact of individual variables on model behavior are beter suited, both theoretically and operationally, to the task of selecting model variables".

Following this line of thought and in an effort to formalize such a subjective process as sensitivity testing, Ford *et al.* [71] have recently developed a technique that makes sensitivity testing less difficult and more automatic.

In testing a given model, the behavior of output variables over the simulation period for variations in the parameters is collected for a great number of runs (100). This vast amount of raw information is then analyzed through

⁷⁰Refer to section (IV-C.1)

formal and informal methods.

In the formal analysis, partial correlation coefficients of an output variable and all inputs are calculated and filtered over time (to isolate "persistent" correlations) to be used in selecting key input variables that appear to have the most influence on the output variable of interest.

In the informal analysis, the partial correlation coefficients are used in combination with the various display options to examine the model's behavior under widely varying input conditions.

These two processes, involving both the model builder and the statistician, allow a model's behavior and weak points to be assessed thoroughly and focus research, careful verification and refinement on the sensitive relationships of the model.

On the other hand, much tedious theoretical work is still being carried out in the field of formal statistical estimation of socio-economic models, and in particular in the use of the Filtering Form of the Maximum Likelihood estimator and indeed, and many researchers throughout the world are continually refining our theoretical understanding of the issues involved.

These efforts however are attempts at stretching out to its limits the practice of parameter estimation as we now conceive it. The past hundred years or so have seen a dramatic break in "normal" trends. We are living in an age

of transition that must be characterized by an equivalent change in the way we look at the world, or a change in "paradigm"⁷¹.

" The ways of looking at things that we have in the past accepted as common sense really do not work under all circumstances and it is very likely that we have reached a period of history when they do not match the types of processes which are going in the world at large..." [117]

The many limitations of traditional model-building and parameter estimation call for a reassessment of the process of model-building itself. Causality in society is a much more "fuzzy" concept than in the physical sciences and yet, we are carrying our physical way of thinking into the field of socio-economic models. Rather than remaining within the framework of deterministic relationships between variables, we should be exploring ways that allow for more flexibility. The investigation of causality through the use of information theoretic concepts such as the mutual information entropy between variables and other less restrictive approaches might well prove to be more suited for the processes at hand.

⁷¹See T. KUHN [116]

REFERENCES

- [1] WARFIELD, J.N., "*Societal systems: Planning, Policy and Complexity* " John Wiley & Sons, New York, 1976.
- [2] BERTALANFFY, L.V., "*General Systems Theory*", George Braziller, New York, 1968.
- [3] WIENER, N., "*Cybernetics*", 2nd. Edition, the MIT Press, Cambridge, Massachussetts, 1961.
- [4] SHANNON, C.E., and W. WEAVER, "*The Mathematical Theory of Communication*", University of Illinois Press, Urbana, 1949.
- [5] FORRESTER, J.W., "*World Dynamics*", Wright Allen, Cambridge, Massachussetts, 1971.
- [6] MEADOWS, D.H., D.L. MEADOWS, J.RANDERS, and W. BEHREN III, "*The Limits to Growth*", Universe Books, New York, 1972.
- [7] MESAROVIC, M. and D. PESTEL, "*Mankind at a Turning Point*", Signet Books, New York, 1974.
- [8] BROWN, T.M., "*Specification and Uses of Econometric Models*", McMillan, St. Martin's Press, London, 1970.
- [9] FORRESTER, J.W., "*Urban Dynamics*", MIT Press, Cambridge, Massachussetts, 1969.
- [10] HOUSE, P.W. and J. McLEOD, "*Large Scale Models for Policy Evaluation*", John Wiley & Sons, New York, 1977.
- [11] SAGE, A.P. "*Methodology for Large-Scale Systems*", McGraw Hill, New York, 1977.

- [12] FORRESTER, J.W. "*Industrial Dynamics*", The MIT Press, Cambridge, Massachusetts, 1961.
- [13] ANSOFF, H.I. and D. SLEVIN, "An Appreciation of Industrial Dynamics", *Management Science*, Vol. 14, (7), March 1968, pp. 383-397.
- [14] SUSSEX GROUP and MEADOWS, et al., "The Limits to Growth Controversy," *Futures*, special issue, Vol. 5, (1), February 1973.
- [15] FORRESTER, J.W., "A Response to Ansoff and Slevin", in "*Collected Papers of Jay W. Forrester*", Wright Allen, Cambridge, Massachusetts, 1975.
- [16] FORRESTER, J.W., "A National Model for Understanding Social and Economic Change", *Simulation*, Vol. 24, (4), April 1975, *Simulation Today*, pp. 125-128.
- [17] SENGE, P.M., "*Testing Estimation Techniques for Social Models*", System Dynamics Group Working Paper D-2199-4, Alfred P. Sloan School of Management, E40-253, MIT, Cambridge, Massachusetts, Nov. 1975.
- [18] SENGE, P.M., "Statistical Estimation of Feedback Models", *Simulation*, Vol. 28, (6), June 1977, pp. 177-184.
- [19] FORRESTER, J.W., "Market-Growth as Influenced by Capital Investment", in "*Collected Papers of Jay Forrester*", Wright Allen, Cambridge, Massachusetts, 1975.
- [20] EYKHOFF, P., "*System Identification: Parameter and State Estimation*", John Wiley & Sons, New York, 1974.
- [21] KALMAN, R.E., "A New Approach to Linear Filtering and Prediction Problems", *Trans ASME, series D, Journal of Basic Engineering*, Vol. 82, 1960, pp. 35-45.

- [22] THEIL, H., "*Principles of Econometrics*", John Wiley & Sons, New York, 1971.
- [23] BOX, G.E.P. and G.M. JENKINS, "*Time Series Analysis: Forecasting and Control*", Holden-Day, San Francisco, 1976.
- [24] MEHRA, R.K., "Identification in Control and Econometrics, Similarities and Differences", *Annals of Economic and Social Measurement*, Vol. 3, (1), Jan. 1974, pp. 21-47.
- [25] ATHANS, M. and G.C. CHON, "Introduction to Stochastic Control Theory and Economic Systems", *Annals of Economic and Social Measurement*, Vol. 1, 1972, pp. 375-384.
- [26] CHOW, G.C., "*Analysis and Control of Dynamic Economic Systems*", Wiley, New York, 1975.
- [27] ATHANS, M., "The Importance of Kalman Filtering Method for Economic Systems", *Annals of Economic and Social Measurement*, Vol. 3, 1974, pp. 49-63.
- [28] ATHANS, M., "The Discrete Time Linear Quadratic Gaussian Stochastic Control Problem", *Annals of Economic and Social Measurement*, Vol. 1, 1972, pp. 449-491.
- [29] BELSEY, D.A., "The Applicability of the Kalman Filter Technique in the Determination of Systematic Parameter Variation", *Annals of Economic and Social Measurement*, Vol. 2, 1973, pp. 531-533.
- [30] DHRYMES, P.J., "*Econometrics: Statistical Foundations and Applications*", Springer-Verlag, New York, 1974.
- [31] JOHNSTON, J., "*Econometric Methods*", McGraw Hill, New York, 2nd. Edition, 1972.
- [32] BRUNDY, J.M. and D.W. JORGENSON, "Consistent and

Efficient Estimation of Simultaneous Equations by Means of Instrumental Variables", in *Frontiers in Econometrics*, P. Zarembka, Editor, Academic Press, New York, 1974.

- [33] MADDALA, G.S., "*Econometrics*", McGraw-Hill, New York, 1977.
- [34] AMEMIYA, T., "Specification Analysis in the Estimation of Parameters of a Simultaneous Equation Model with Autoregressive Residuals", *Econometrica*, Vol. 35, 1966, pp. 283-306.
- [35] FAIR, R.C., "Efficient Estimation of Simultaneous Equations with Autoregressive Residuals by Instrumental Variables", *Review of Economics and Statistics*, Vol. 54, 1972, pp. 444-449.
- [36] GODFREY, L.G., "Testing for Serial Correlation in Dynamic Simultaneous Equation Models", *Econometrica*, Vol. 44, 1976, p. 1077
- [37] HATANAKA, M., "Several Efficient Two Stag Estimators for the Dynamic Simultaneous Equations Model with Autoregressive Errors", *Journal of Econometrics*, Vol. 4, 1976, pp. 184-204.
- [38] HENDRY, H.F., "Maximum Likelihood Estimation of Systems of Simultaneous Regression Equations with Errors Generated by a Vector Autoregressive Process", *International Economic Review*, Vol. 12, 1971, pp. 257-272.
- [39] GOLDBERGER, A.S. and O.D. DUCAN, Editors, "*Structural Equation Models in the Social Sciences*", Seminar Press, New York, 1973.
- [40] WILEY, D.E., "The Identification Problem for Structural Equation Models with Unmeasured Variables", in "*Structural Equation Models in the Social Sciences*", A.S. Goldberger and O.D. Duncan, Editors, Seminar Press, New

York, 1973, pp. 69-83.

- [41] JORESKOG, K.G., "A General Method for the Analysis of Covariance Structures", *Biometrika*, 1970, (57), pp. 239-251.
- [42] JORESKOG, K.G., "A General Method for Estimating a Linear Structural Equation System", in "*Structural Equation Systems in the Social Sciences*", A.S. GOLDBERGER and O.D. DUNCAN, Editors, Seminar Press, New York, 1973.
- [43] BASMANN, R.L., "A Note on the Exact Finite Sample Frequency Functions of Generalized Classical Linear Estimators in Two Leading Overidentified Cases", *Journal of the American Statistical Association*, Vol. 56, 1961, pp. 619-636.
- [44] BASMANN, R.L., "A Note on the Exact Finite Sample Frequency Functions of Generalized Classical Linear Estimators in a Leading Three Equation Journal of the American Statistical Association, Vol. 58, 1961.
- [45] NAGAR, A.L., "Double k-Class Estimators of Parameters in Simultaneous Equations and their Small-Sample Properties", *International Economic Review*, Vol. 3, May 1962, pp. 168-188.
- [46] SUMMERS, R., "A Capital Intensive Approach to the Small-Sample Properties of Various Simultaneous Equations Estimators", *Econometrica*, Vol. 35, 1965, pp. 1-41.
- [47] WAGNER, H.M., "A Monte-Carlo Study of Estimates of Simultaneous Linear Structural Equation", *Econometrica*, Vol. 26, 1958, pp. 117-133.
- [48] NAGAR, A.L., "A Monte-Carlo Study of Alternative Simultaneous Equations Estimators", *Econometrica*, Vol. 28, 1960, pp. 573-590.
- [49] CRAGG, J.G., "On the Relative Small-Sample

Properties of several Structural Equation Estimators", *Econometrica*, Vol.35, 1967, p. 573

- [50] BASMANN, R.L., et al., "finite Sample Distributions Associated with Stochastic Difference Equations - Some Experimental Evidence", *Econometrica*, Vol. 42, 1974, pp. 825-840.
- [51] MADDALA, G.S., "Some Small-Sample Evidence on Tests of Significance in Simultaneous Equations Models", *Econometrica*, Vol. 42, 1974, pp. 841-852.
- [52] MARIANO, R.S., "Finite Sample Properties of Instrumental Variable Estimators of Structural Coefficients", *Econometrica*, Vol. 45, 1977, pp. 487-498.
- [53] GOLDFELD, S.M., and R.E. QUANDT, "Nonlinear Simultaneous Equations Estimation and Prediction", *International Economic Review*, Vol. 9, 1968, pp. 113-136.
- [54] EISENPRESS, H. and J. GREENSTDT, "The Estimation of Nonlinear Econometric Systems", *Econometrica*, Vol. 34, 1966, pp. 851-861.
- [55] AMEMIYA, T., "The Nonlinear Two-Stage Least Squares Estimator", *Journal of Econometrics*, Vol. 2, 1974, pp. 105-110.
- [56] AMEMIYA, T., "The Nonlinear Maximum Likelihood Estimator and the Modified Nonlinear Two-Stage Least Squares Estimator", *Journal of Econometrics*, Vol. 3, 1975, pp. 375-386.
- [57] GALLANT, A.R., "3SLS Estimation for a System of Simultaneous Nonlinear Implicit Equations", *Journal of Econometrics*, Vol. 5, 1977, pp. 71-88.
- [58] HAUSMANN, J.A., "An Instrumental Variables Approach to Full-Information Estimation for

Linear and Certain Nonlinear Econometric models", *Econometrica*, Vol. 45, 1975, pp. 727-738.

- [59] JORGENSON, D.W., and J. LAFFONT, "Efficient Estimation of Nonlinear Simultaneous Equations with Additive Disturbances", *Annals of Economic and Social Measurement*, Vol. 3, 1974, pp 615-616
- [60] BERNDT, E.K., et al., "Estimation and Inference in Nonlinear Structural Models", *Annals of Economic and Social Measurement*, Vol. 3, Oct. 1974, pp. 653-666.
- [61] GOLDFELD, S.M., and R.E. QUANDT, "*Nonlinear Methods in Econometrics*", North-Holland, New York, 1972.
- [62] CHOW, G.C., "On the Computation of Full-Information Maximum Likelihood Estimates for Nonlinear Equation Systems", *Review of Economics and Statistics*, Feb. 1973, pp. 104-109.
- [63] BELSEY, D.A., "Estimation of Systems of Simultaneous Equations and Computational Specifications of GREMLIN", *Annals of Economic and Social Measurement*, Vol. 3, Oct. 1974, pp. 551-614.
- [64] JORESKOG, K.G., and D. SORBOM, "*LISREL III, Estimation of Linear Structural Equation Systems by Maximum Likelihood Methods. - A FORTRAN IV Program*", National Educational Resources, Inc., Chicago, Illinois, 1976.
- [65] FLETCHER, R. and M.J.D. POWELL, "A Rapidly Convergent Descent Method for Minimization", *The Computer Journal*, Vol. 6, 1967, pp. 163-168.
- [66] GRUVAEUS, G.T., and K.G. JORESKOG, "A Computer Program for Minimizing a Function of Several Variables", *Research Bulletin*, 70 - 14, Educational Testing Service, Princeton, New

Jersey, 1970.

- [67] ASTROM, K.J., and P. EYKHOFF, "System Identification - A Survey", *Automatica*, Vol. 7, 1971, pp. 123-162.
- [68] JAZWINSKY, A.H., "*Stochastic Processes and Filtering Theory*," Academic Press, New York, 1970.
- [69] NAHI, N.E., "*Estimation Theory and Applications*", John Wiley and Sons, New York, 1969.
- [70] MASS, N.J. and P.M. SENGE, "Alternative Tests for the Selection of Model Variables", *IEEE Transactions on Systems, Man and Cybernetics*, Vol. SMC-8, (6), June 1978, pp. 450-460.
- [71] FORD, A., G.H. MOORE and M.D. McKAY, "Sensitivity Analysis of Large Computer Models: A Case Study of the COAL2 National Energy Model", Rough Draft, Los Alamos Scientific Laboratory, Oct. 1978
- [72] ZABORSKY, J., "An Information Theory Viewpoint for the General Identification Problem", Correspondence, *IEEE Transactions on Automatic Control*, Vol. AC-11, (1), Jan. 1966, pp. 130-131.
- [73] CAINES, P.M., "Stationary Linear and Nonlinear System Identification and Predictor Set Completeness", *IEEE Transactions on Automatic Control*, Vol. AC-23, (4), Aug. 1978, pp. 583-594.
- [74] KULLBACK, S., "*Information Theory and Statistics*", John Wiley and Sons, New York, 1959.
- [75] KAILATH, T., "The Divergence and Bhattacharyya Distance Measures in Signal Selection", *IEEE Transactions on Automatic Control*, Vol. COM-15, (1), Feb. 1967, pp. 52-60

- [76] SCHWEPPE, F.C., "On the Bhattacharyya Distance and the Divergence Between Gaussian Processes", *"Information and Control"*, Vol. 11, (4), Oct. 1967, pp. 373-395.
- [77] AKAIKE, H., "Canonical Correlation Analysis of Time Series and the Use of an Information Criterion", in *"System Identification: Advances and Case Studies"*, R.K. MEHRA and D.G. LAINIOTIS, Editors, Academic Press, New York, 1976.
- [78] AKAIKE, H., "A New Look at the Statistical Model Identification", *IEEE Transactions on Automatic Control*, Vol. AC-19, (6), Dec. 1974, pp. 716-723.
- [79] RISSANEN, J., "Minimax Entropy Estimation of Models for Vector Processes", in *"System Identification: Advances and Case Studies"*, R.K. MEHRA and D.G. LAINIOTIS, Editors, Academic Press, New York, 1976.
- [80] TSE, E., "A Quantitative Measure of Identifiability", *IEEE Transactions on Systems, Man and Cybernetics*, Vol. MSC-8, (1), Jan. 1978, pp. 1-8.
- [81] FISHER, F.M., *"The Identification Problem in Econometrics"*, McGraw Hill, New York, 1966.
- [82] TSE, E. and J.J. ANTON, "On the Identifiability of Parameters", *IEEE Transactions on Automatic Control*, Vol. AC-17, (5), Oct. 1972, pp. 637-645.
- [83] BLACKMANN, M., *"Introduction to State Variable Analysis"*, McMillan Press, 1977.
- [84] MEHRA, R.K., and J.S. TYLER, "Case Studies in Aircraft Parameter Identification", 1973 IFAC Symposium on Identification and System Parameter Estimation, The Hague, Netherlands.
- [85] PETERSON, D. W., *"GPSIE: General Purpose System*

Identifier and Evaluator", User's Manual, Draft, 1976.

- [86] SCHWEPPE, F. C., "*Uncertain Dynamic Systems*", Prentice Hall, Englewood Cliffs, New Jersey, 1973.
- [87] SCHWEPPE, F. C., "*Evaluation of Likelihood Functions for Gaussian Signals*", *IEEE Transactions on Information Theory*, Vol. IT-11, No. 1, Jan. 1965.
- [88] IMSL library, Sixth Edition, Customer Relations, Sixth Floor, GNB Building, 7500 Bellaire Blvd., Houston, TX 77036.
- [89] CHENIER, J.P., Editor, "*TIME SERIES PROCESSOR*", Computing Services, University of Alberta, Edmonton, Alberta, 1978.
- [90] MORGENSTERN, O., "*On the Accuracy of Economic Observations*", Princeton University, Princeton, Mass., 1963.
- [91] WOLD, H.O., Editor, "*Econometric Model Building*", North Holland, 1964.
- [92] FISHER, F.M., "A Correspondence Principle for Simultaneous Equations Models", *Econometrica*, Vol. 38, (1), Jan. 1970.
- [93] SENGE, P.M., "*A Reexamination of the Foundations for Simultaneous Equations Systems*", System Dynamics Group Working Paper D-1748, Alfred P. Sloan School of Management, E40-253, MIT, Cambridge, Mass., Dec. 1973.
- [94] MORECROFT, J. D. W., "*A comparison of OLS and IV Estimation of a Dynamic Growth Model*", Paper D-2729, Sloan School of Management, MIT, 1977.
- [95] ASTROM, K. S., and S. WENMARK, "*Numerical Identification of Stationary Time-Series*",

Sixth International Instruments and
Measurements Congress, Stockholm, Sept. 1964.

- [96] KASHYAP, R. L., *Maximum Likelihood Identification of Stochastic Linear Dynamic Systems*", IEEE Transactions on Automatic Control, Vol. AC-15, Feb. 1970, pp. 25-34.
- [97] MEHRA, R. K., "Identification of Stochastic Linear Dynamic Systems Using Kalman Filter Representation", AIAA Journal, Vol. 9, No. 1, Jan. 1971, pp. 28-31.
- [98] ASTROM, K.J. and T. BOHLIN, "Numerical Identification of Linear Dynamic Systems from Normal Operating Records", in "Theory of Self-Adaptive Control Systems", P. HAMMOND, Editor, Plenum Press, New York, 1965.
- [99] EATON, J., "Identification for Control Purposes, IEEE Winter Meeting, New York, 1967, (IEEE International Convention Record, Part 3, Automatic Control.
- [100] CAINES, P.E., and J. RISSANEN, "Maximum Likelihood Estimation of Parameters in Multivariable Gaussian Stochastic Processes", IEEE Transactions on Information Theory, Vol. IT-20, (1), Jan. 1974, Correspondence, pp. 102-104
- [101] WOO, K.T., "Maximum Likelihood Identification of Noisy Systems", 2nd IFAC Symposium on Identification and Process Parameter Estimation, Prague, 1970.
- [102] LJUNG, L., "On the Consistency of Prediction Error Identification Methods", in *System Identification: Advances and Cases Studies*", R.K. MEHRA and D.G. LAINIOTIS, Editors, Academic Press, New York, 1976.
- [103] OLLSON, G., "Modeling and Identification of a Nuclear Reactor", in "System Identification: Advances and Case Studies", R. K. MEHRA and D. G. LAINIOTIS, Editors, Academic Press, New

York, 1976.

- [104] LEONDES, C.T., Editor, "*Theory and Applications of Kalman Filtering*", AGARDO-graphs 139, ***, 139.
- [105] ASTROM, K., "*Introduction to Stochastic Control Theory*", Academic Press, New York, 1970.
- [106] LEE, E.B. and L. MARKUS, "*Foundatios of Optimal Control Theory*", Wiley, New York, 1967.
- [107] SAGE, A.P. and J.L. MELSА, "*Estiamtion Theory with Applications to Communications and Control*", McGraw Hill, New York, 1971.
- [108] MEHRA, R.K., "A Comparison of Several Nonlinear Filters for Radar Target Tracking", **IEEE Transactions on Automatic Control**, Vol. AC-16, (4), Aug. 1971.
- [109] JAZWINSKY, A.H., "Filtering for Nonlinear Dynamical Systems", **IEEE Transactions on Automatic Control**, Vol. AC-11, 1966, p. 765
- [110] DETCHMENDY, D.M. and R. SRIDHAR, "Sequential Estimation of States and Parameters in Noisy Nonlinear Dynamical Systems", **Journal of Basic Engineering**, Vol. 88, Series D, 1966, pp. 209-264.
- [111] KAILATH, T., "*An Innovations Approach to Least Squares Estimation: Part I*", **IEEE Transactions on Automatic Control**, Vol. AC-13, 1968, pp. 646-655.
- [112] MEHRA, R.K. and N.K. GUPTA, "Computational Aspects of Maximum Likelihood Estiamtion and Reduction in Sensitivity Function Calculations", **IEEE Transactions on Automatic Control**, Vol. AC-19, (6), Dec. 1974, pp. 716-723.
- [113] FMFP: Application Program-System/360 Scientific

Subroutine Package (360A-CM-03X), Version 3,
1968.

- [114] NISHIMURA, T., "Modeling Errors in Kalman Filters", in *"Theory and Applications of Kalman Filtering"*, C.T. LEONDES, Editor, AGARDO-graphs 139, 1970.

- [115] LJUNG, L., "Asymptotic Behavior of the Extended Kalman Filter as a Parameter Estimator for Linear Systems", *IEEE Transactions on Automatic Control*, Vol. AC-24, No.1, Feb. 1979, pp. 36-50.

- [116] KUHN, T. S., "The Structure of Scientific Revolutins", 2nd. Edition, The University of Chicago Press, 1970.

- [117] WADDINGTON, C. H., "Tools for Thought", Jonathan Cape, London, 1977.

B30304